GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 17:59:42 ; Search time 2777 Seconds
(without alignments)
2419.514 Million cell updates/sec

Title: US-09-765-231A-58
Perfect score: 225
Sequence: 1 tgatggtaagttgtttcagg.....attaggaatttttttttttt 225

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	196	87.1	341	9	AA502552	AA502552 ng62e06.s
	2	117	52.0	324	9	AI247782	AI247782 qh61a07.x
	3	117	52.0	362	9	AA682512	AA682512 zi19a01.s
	4	48.8	21.7	928	29	CNS00DKY	AL071865 Drosophil
c	5	47	20.9	1131	13	BX356147	BX356147 BX356147
c	6	44.6	19.8	1201	13	BX324729	BX324729 BX324729
c	7	44.2	19.6	1108	13	BX418757	BX418757 BX418757
	8	43.8	19.5	1155	9	AL514851	AL514851 AL514851
	9	43	19.1	773	28	AQ781761	AQ781761 HS_3122_A
c	10	42.8	19.0	1091	13	BX424950	BX424950 BX424950
	11	42.4	18.8	1121	13	BX338325	BX338325 BX338325
c	12	42.2	18.8	542	13	BU722265	BU722265 SJMAFC03
c	13	42.2	18.8	1056	13	BX415058	BX415058 BX415058
c	14	42	18.7	1201	13	BX462207	BX462207 BX462207
c	15	41.8	18.6	964	29	CNS058MA	AL326107 Tetraodon
c	16	41.2	18.3	1101	29	CNS00D9O	AL060402 Drosophil
	17	41	18.2	290	10	AW504318	AW504318 UI-HF-BN0
c	18	41	18.2	451	14	CB051834	CB051834 NISC_g103
	19	41	18.2	1562	11	BC022863	BC022863 Homo sapi
c	20	40.6	18.0	595	28	AZ523166	AZ523166 216PbB05
c	21	40.6	18.0	957	29	CNS015W7	AL105985 Drosophil
c	22	40.4	18.0	516	13	BX561821	BX561821 BX561821
c	23	40.4	18.0	525	10	AW381028	AW381028 CM1-HT028
c	24	40.4	18.0	999	13	BX380865	BX380865 BX380865
	25	40.4	18.0	1201	13	BX385531	BX385531 BX385531
c	26	40	17.8	475	10	BE325739	BE325739 NF055G02S
	27	40	17.8	611	29	CE160147	CE160147 tigr-gss-
	28	40	17.8	945	13	BX418213	BX418213 BX418213
c	29	40	17.8	1081	28	CC247576	CC247576 CH261-111
c	30	40	17.8	1098	13	BX377526	BX377526 BX377526
	31	40	17.8	1201	13	BX458169	BX458169 BX458169
	32	39.8	17.7	208	9	AU071524	AU071524 AU071524
	33	39.8	17.7	638	14	CD649691	CD649691 CvGil0039
c	34	39.8	17.7	647	28	B83740	B83740 CpG0090B Cp
c	35	39.8	17.7	800	14	CB971606	CB971606 CAB10005_
c	36	39.8	17.7	890	14	CB756780	CB756780 AGENCOURT
	37	39.8	17.7	1201	13	BX421216	BX421216 BX421216
c	38	39.6	17.6	358	14	CD801440	CD801440 EST 9349
	39	39.6	17.6	360	9	AL750985	AL750985 AL750985
	40	39.6	17.6	409	13	BQ451684	BQ451684 PfESTOab0
	41	39.6	17.6	524	28	AZ055107	AZ055107 RPCI-23-4
	42	39.6	17.6	941	28	BH133309	BH133309 ENTOQ10TF
c	43	39.6	17.6	1195	13	BX355698	BX355698 BX355698
	44	39.4	17.5	189	12	BJ351968	BJ351968 BJ351968
	45	39.4	17.5	985	9	AL520226	AL520226 AL520226

ALIGNMENTS

RESULT 1
AA502552
LOCUS AA502552 341 bp mRNA linear EST 19-AUG-1997
DEFINITION ng62e06.s1 NCI_CGAP_Lip2 Homo sapiens cDNA clone IMAGE:939394, mRNA sequence.
ACCESSION AA502552
VERSION AA502552.1 GI:2237519
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 341)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: David B. Krizman, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

```
            Insert Length: 1183    Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 310.
FEATURES              Location/Qualifiers
     source           1. .341
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:939394"
                      /tissue_type="liposarcoma"
                      /lab_host="DH10B"
                      /clone_lib="NCI_CGAP_Lip2"
                      /note="Vector: pAMP10; mRNA made from liposarcoma, cDNA
                      made by oligo-dT priming. Non- directionally cloned.
                      Size-selected on agarose gel, average insert  size 600 bp.
                      Reference:  Krizman et al. (1996) Cancer Research
                      56:5380-5383."
ORIGIN

  Query Match             87.1%;  Score 196;  DB 9;  Length 341;
  Best Local Similarity   99.5%;  Pred. No. 5.9e-28;
  Matches  207;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;

Qy         19 GGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCA 78
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          8 GGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCA 67

Qy         79 GAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAAC 138
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         68 GAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAAC 127

Qy        139 AAAAC-AAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTG 197
              ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        128 AAAACAAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTG 187

Qy        198 CTGAATCAATTAGGAATTTTTTTTTTTT 225
              ||||||||||||||||||||||||||||
Db        188 CTGAATCAATTAGGAATTTTTTTTTTTT 215


RESULT 2
AI247782
LOCUS       AI247782                324 bp    mRNA    linear   EST 01-DEC-1998
DEFINITION  qh61a07.x1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:1849140 3', mRNA sequence.
ACCESSION   AI247782
VERSION     AI247782.1  GI:3843179
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 324)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert Length: 821    Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 317.
FEATURES              Location/Qualifiers
     source           1. .324
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:1849140"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      This is a subtracted version of the original Soares fetal
                      liver spleen 1NFLS library.  1st strand cDNA was primed
                      with a Pac I - oligo(dT) primer [5'
                      AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
                      (Pharmacia), digested with Pac I and cloned into the Pac I
                      and Eco RI sites of the modified pT7T3 vector.  Library
                      went through one round of normalization. Library
                      constructed by Bento Soares and M.Fatima Bonaldo."
ORIGIN

  Query Match             52.0%;  Score 117;  DB 9;  Length 324;
  Best Local Similarity   100.0%;  Pred. No. 7e-13;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        109 AAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAAACAACCTTTTTAAGAGTTG 168
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAAACAACCTTTTTAAGAGTTG 60

Qy        169 ATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 117


RESULT 3
AA682512
LOCUS       AA682512                362 bp    mRNA    linear   EST 19-DEC-1997
DEFINITION  zi19a01.s1 Soares_fetal_liver_spleen_1NFLS_S1 Homo sapiens cDNA
            clone IMAGE:431208 3', mRNA sequence.
ACCESSION   AA682512
VERSION     AA682512.1  GI:2669793
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 362)
  AUTHORS   Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M.,
            Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F.,
            Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
  TITLE     WashU-NCI human EST Project
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 308.
FEATURES              Location/Qualifiers
     source           1. .362
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="IMAGE:431208"
                      /sex="male"
                      /dev_stage="20 week-post conception fetus"
                      /lab_host="DH10B (ampicillin resistant)"
                      /clone_lib="Soares_fetal_liver_spleen_1NFLS_S1"
                      /note="Organ: Liver and Spleen; Vector: pT7T3D (Pharmacia)
                      with a modified polylinker; Site_1: Pac I; Site_2: Eco RI;
                      This is a subtracted version of the original Soares fetal
                      liver spleen 1NFLS library.  1st strand cDNA was primed
                      with a Pac I - oligo(dT) primer [5'
                      AACTGGAAGAATTAATTAAAGATCTTTTTTTTTTTTTTTTT 3'],
                      double-stranded cDNA was ligated to Eco RI adaptors
```

(Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match 52.0%; Score 117; DB 9; Length 362;
Best Local Similarity 100.0%; Pred. No. 6.5e-13;
Matches 117; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy        109 AAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTG 168
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 AAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTG 60

Qy        169 ATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 117
```

RESULT 4
CNS00DKY
LOCUS CNS00DKY 928 bp DNA linear GSS 04-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC # BACR27A24 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL071865
VERSION AL071865.1 GI:4948170
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 928)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES Location/Qualifiers
source 1. .928
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR27A24"
/clone_lib="RPCI-98"
/note="end : T7"

ORIGIN

Query Match 21.7%; Score 48.8; DB 29; Length 928;
Best Local Similarity 31.4%; Pred. No. 3.5;
Matches 64; Conservative 59; Mismatches 81; Indels 0; Gaps 0;

```
Qy         22 ATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAA 81
              | ||| ||| :||: || : | | :| | | | | ||  |:|: | : ||
Db        678 AAAAATTTTAWAAWTAAAAWAAAAATTTAWTTTTTTTTTTTTTTTWTATWTWAWATAWAAA 737

Qy         82 GAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAA 141
```

```
              | :::::: |:| : : :    ::: :::  :|   :       ::::::  ||| |||
Db        738 TATWWWWWWATWTWDGKNWNNNAWWWWWWWWWWAWWWWAAAWWWWWWWWWWAAAAAAAA 797

Qy        142 ACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGA 201
              | |||:||| ||  :::::::::::::| :: :  :  :::: : :|  :  | |  :||
Db        798 AAAAAWAAAAAAAWDDDDDDWDDDWWKAAKKKKKKKKKKKKKKKKKBGKKKTKTKTTTKGA 857

Qy        202 ATCAATTAGGAATTTTTTTTTTTT 225
              ::   ||     ||||||||||||
Db        858 RWWTTTTTTTTTTTTTTTTTTT 881
```

RESULT 5
BX356147/c
LOCUS BX356147 1131 bp mRNA linear EST 05-MAY-2003
DEFINITION BX356147 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DI008YB06 5-PRIME, mRNA sequence.
ACCESSION BX356147
VERSION BX356147.1 GI:30378029
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1131)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 9728.r For more information about this cluster, see http://www.genoscope.cns.fr/cgi-bin/cluster.cgi?seq=CS0DI008DA03QP1&cluster=9728.r. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DI008DA03QP1.
FEATURES Location/Qualifiers
source 1. .1131
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DI008YB06"
/tissue_type="PLACENTA COT 25-NORMALIZED"
/clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
/note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match 20.9%; Score 47; DB 13; Length 1131;
Best Local Similarity 37.9%; Pred. No. 6.8;
Matches 77; Conservative 39; Mismatches 87; Indels 0; Gaps 0;

```
Qy         23 TAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAG 82
              |: : |||     :::: |  |   |  | |  |   ||| ||||   |: : ||
Db        559 THCMTTTTTTYCCMWWWAAAAAATTTTWAAAAAAACCCCCATTTTTTTTTTTTWWWWAAA 500

Qy         83 AATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAA 142
              |: : ::: ||:| | |   |||| ::||   || : :::|||| :|||| :|||   :
Db        499 AMCMMWWWWTTWCCCCCCYTTTTTMMMAAAAAACCBSKBSCCCCBWAAAAWWWAACCCCM 440

Qy        143 CAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAA 202
              :|||||| ||       || |   |   :: |  ||| | |   | | :  : ::
Db        439 MAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWWWWTTTTTTTTTTTTTTTTTWTTWWWWAWW 380

Qy        203 TCAATTAGGAATTTTTTTTTTT 225
                ||:::  ::::|::|||||||
Db        379 AAAAWWWWWWWWWTWWTTTTTTT 357
```

```
RESULT 6
BX324729/c
LOCUS       BX324729               1201 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX324729 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI037YJ05 5-PRIME, mRNA sequence.
ACCESSION   BX324729
VERSION     BX324729.1  GI:30338394
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3281.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0AI037CE03QP1&cluster=3281.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0AI037CE03QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI037YJ05"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match            19.8%;  Score 44.6;  DB 13;  Length 1201;
  Best Local Similarity  36.3%;  Pred. No. 19;
  Matches   81;  Conservative   41;  Mismatches  101;  Indels    0;  Gaps    0;

Qy          3 ATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTAT 62
              | ||  |   |::||: | ||: |:  ||||| |||   : | :: | : | |:    |
Db        695 AAGGRGAWRGGKKTCMAGGATMTAVGGTGAAAAAAAAMSYCAARSSCTRGGTTRAAAAAA 636

Qy         63 ATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTC 122
              |     | |     | ||  |||  :|     |   |         ||| || ||
Db        635 AAAAAATAAATAAAAAAAAAATATAAAWAAAAAAAAAAAAAAAAAAAAAAAAAATAGGCCCCT 576

Qy        123 CCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTT 182
              |   | :|||  ||| |||| :|||||| ::  ::::::: : :: ::   |  | :::|
Db        575 CAAAATWAAAATAAAAAAAAAAMAAAAAAAWWWWWWWWWWWWWWWWTWWTWTTTTWWWT 516

Qy        183 GATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              | | |  | | |  | ::: :::::  :::|||||||||||
Db        515 TTTWTTTTTTTTTTTTTTWWWWWWWWWWWWWWTTTTTTTTTTT 473


RESULT 7
BX418757/c
LOCUS       BX418757               1108 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX418757 Homo sapiens FETAL BRAIN Homo sapiens cDNA clone
            CS0DF009YG18 5-PRIME, mRNA sequence.
ACCESSION   BX418757
VERSION     BX418757.1  GI:30769508
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1108)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DF009BD09QP1.
FEATURES             Location/Qualifiers
     source          1. .1108
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DF009YG18"
                     /tissue_type="FETAL BRAIN"
                     /dev_stage="fetal"
                     /clone_lib="Homo sapiens FETAL BRAIN"
                     /note="Organ: brain; Vector: pCMVSPORT_6; 1st strand cDNA
                     was primed with a NotI-oligo(dT) primer. Five prime end
                     enriched, double-strand cDNA was digested with Not I and
                     cloned into the Not I and EcoRV sites of the pCMVSPORT 6
                     vector. Library was not normalized."
ORIGIN

  Query Match            19.6%;  Score 44.2;  DB 13;  Length 1108;
  Best Local Similarity  32.7%;  Pred. No. 23;
  Matches   64;  Conservative   49;  Mismatches  83;  Indels    0;  Gaps    0;

Qy         30 TGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAGAATATTT 89
              | : |:| |||   |   |  ||   ||  ||||    : ||| |||    : || |||
Db        722 TCMCAYACATTTCAATAATAAATATCATAWTATAACATWATTATCTTTTTCMTAAAATTC 663

Qy         90 AGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAA 149
              |  |   |  | ||||| ||      :   :   ||:|||  ||| ||| |:|:::|
Db        662 ATATATNCCAAAATTTTTTAATATATAATWMCWMAAAAAWAAAAAAAAAAAACCMAMMMA 603

Qy        150 ACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTA 209
              :|  |:     :: :   :   :: ::: :::  ::: ::::::|: :: ::: :::::
Db        602 MCCCCMMMMAMMMMMCMMMMMCCYYYYYYYYYYYYYYYYYYYYYYCYYYYWWWWAWWWWW 543

Qy        210 GGAATTTTTTTTTTTT 225
                ::: |:||||||||
Db        542 WWWWWATWTTTTTTTT 527


RESULT 8
AL514851
LOCUS       AL514851               1155 bp    mRNA    linear   EST 08-MAY-2003
DEFINITION  AL514851 Homo sapiens NEUROBLASTOMA Homo sapiens cDNA clone
            CL0BB014ZC07 3-PRIME, mRNA sequence.
ACCESSION   AL514851
VERSION     AL514851.2  GI:30464736
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1155)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     On Feb 13, 2001 this sequence version replaced gi:12778344.
            Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
```

```
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 3301.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CL0BB014ZC07FP1&cluster=3301.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CL0BB014ZC07FP1.
FEATURES              Location/Qualifiers
     source           1. .1155
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CL0BB014ZC07"
                      /tissue_type="NEUROBLASTOMA"
                      /clone_lib="Homo sapiens NEUROBLASTOMA"
                      /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                      with a NotI-oligo(dT) primer. Five prime end enriched,
                      double-strand cDNA was digested with Not I and cloned into
                      the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                      Library was not normalized."
ORIGIN

  Query Match             19.5%;  Score 43.8;  DB 9;  Length 1155;
  Best Local Similarity   43.0%;  Pred. No. 27;
  Matches   65;  Conservative   26;  Mismatches   60;  Indels    0;  Gaps    0;

Qy          75 TTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACG 134
               |:| |:| ||  |||  |: | |   || :|||    | | |:  |||||   :||
Db         868 TYCSGRAAAAATTTTTTTYCCCCSGGGGGYTTTTTTYYCCCCCCYYCCCCCCCCCMAAAA 927

Qy         135 AAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCT 194
               ||| |||| ||||||| ||  |||:: | |    :   |: : |:    : :: : |  :
Db         928 AAAAAAAAAAAAAAAAAAAAAWTTTYYCAAAAAMCMCCCCYYYCCMMMMAMMHYYYTTTWW 987

Qy         195 CTGCTGAATCAATTAGGAATTTTTTTTTTT 225
               :: ::  : :||    :::|||||||||||
Db         988 YWYYYCCMAMAAMSSCVWWTTTTTTTTTTT 1018


RESULT 9
AQ781761
LOCUS       AQ781761                 773 bp    DNA     linear   GSS 02-AUG-1999
DEFINITION  HS_3122_A2_D08_T7C CIT Approved Human Genomic Sperm Library D Homo
            sapiens genomic clone Plate=3122 Col=16 Row=G, genomic survey
            sequence.
ACCESSION   AQ781761
VERSION     AQ781761.1  GI:5684721
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 773)
  AUTHORS   Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
  TITLE     Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
  MEDLINE   99380589
   PUBMED   10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones may be purchased from Research Genetics (info@resgen.com).
            BAC end Web Server: http://www.htsc.washington.edu
            Plate: 3122  row: G  column: 16
            Seq primer: T7
            Class: BAC ends
            High quality sequence stop: 773.
FEATURES              Location/Qualifiers
     source           1. .773
                      /organism="Homo sapiens"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:9606"
                      /clone="Plate=3122 Col=16 Row=G"
                      /sex="male"
                      /clone_lib="CIT Approved Human Genomic Sperm Library D"
                      /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                      E-Coli DH10B"
ORIGIN

  Query Match             19.1%;  Score 43;  DB 28;  Length 773;
  Best Local Similarity   59.3%;  Pred. No. 51;
  Matches   73;  Conservative    0;  Mismatches   50;  Indels    0;  Gaps    0;

Qy         103 TTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAA 162
               ||||| |||      ||   |||||||||||  |  | |||| |||||||  | || |||| |
Db         188 TTTTTAAAAAAGGGGCCCCCCCCCAAAAAAAGGGAGAAAAAAAAAAAAAAACCCCTTTTTA 247

Qy         163 GAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTT 222
               |       |   |  ||| || |   ||| |     ||  ||  |  ||||||||||
Db         248 AAAAACCCCCCCCTTTTTTAATTTTTTTCCACAAAAAAAAAAAAAAAAAAATTTTTTTTT 307

Qy         223 TTT 225
               |||
Db         308 TAT 310


RESULT 10
BX424950/c
LOCUS       BX424950                 1091 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX424950 Homo sapiens PLACENTA Homo sapiens cDNA clone CL0BA005ZA01
            3-PRIME, mRNA sequence.
ACCESSION   BX424950
VERSION     BX424950.1  GI:30770416
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1091)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 6403.r For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CL0BA005ZA01FP1&cluster=6403.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CL0BA005ZA01FP1.
FEATURES              Location/Qualifiers
     source           1. .1091
                      /organism="Homo sapiens"
                      /mol_type="mRNA"
                      /db_xref="taxon:9606"
                      /clone="CL0BA005ZA01"
                      /tissue_type="PLACENTA"
                      /clone_lib="Homo sapiens PLACENTA"
                      /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                      with a NotI-oligo(dT) primer. Five prime end enriched,
                      double-strand cDNA was digested with Not I and cloned into
                      the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                      Library was not normalized."
```

```
ORIGIN

  Query Match             19.0%;  Score 42.8;  DB 13;  Length 1091;
  Best Local Similarity   21.1%;  Pred. No. 44;
  Matches   46;  Conservative   87;  Mismatches   85;  Indels    0;  Gaps    0;

Qy          8 AAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGG 67
              || ::::::  |: : ||: :::    ||||  |:: :: |  |:: : :: : :::
Db        892 AAADKKKKKKKGKDDAAAVGKKKTTKKTAAAAAAKKKKKKTWWADDKKKKKKKKDDAAA 833

Qy         68 TTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCA 127
               :::   :  | || || |    ::::   :   :::::: ||  :: : :: :::::::
Db        832 AKKKKKKDAAAAAAAAAAAAAANKKKKKKKKKKKKTKKKDDWAAAYYYYYYYYYYYYYYYHH 773

Qy        128 AAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCT 187
              |||||  ||| |||| |:: ::: |  :     :: |     :  :: ::: :::  :::
Db        772 AAAAAAAAAAAAAAAAAWWNHHWTACAMMAMAAMMAAAAAYYYYYYYYYYYYYYYYYYYY 713

Qy        188 GCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTT 225
              ::::::::  : |:|   ||     :::::|||||||
Db        712 YYYYYYYYYWWWAAWTTTTTTTTTYYYYYYYTTTTTT 675


RESULT 11
BX338325
LOCUS       BX338325                1121 bp    mRNA    linear   EST 02-MAY-2003
DEFINITION  BX338325 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
            clone CS0DI058YK15 5-PRIME, mRNA sequence.
ACCESSION   BX338325
VERSION     BX338325.1  GI:30337694
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1121)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 5962.f For
            more information about this cluster, see
            http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DI058AF08QP1&cluster=5962.f. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DI058AF08QP1.
FEATURES             Location/Qualifiers
     source          1. .1121
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI058YK15"
                     /tissue_type="PLACENTA COT 25-NORMALIZED"
                     /clone_lib="Homo sapiens PLACENTA COT 25-NORMALIZED"
                     /note="1st strand cDNA was primed with a NotI-oligo(dT)
                     primer. Five prime end enriched, double-strand cDNA was
                     digested with Not I and cloned into the Not I and EcoR V
                     sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

  Query Match             18.8%;  Score 42.4;  DB 13;  Length 1121;
  Best Local Similarity   39.9%;  Pred. No. 51;
  Matches   59;  Conservative   32;  Mismatches   57;  Indels    0;  Gaps    0;

Qy         78 AGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAA 137
              | || || |        ::: ::  :::::::: |: : :: :::::: | ||| |||
Db        738 AAAAAAAAAAAANANAAMMMMMMCMYYYYHHHHCYCYYYYYYYYYYYYYTATAAAACAAA 797

Qy        138 CAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTG 197
              :|||| :||||| || |||:||| | | ||  || :|||| | |  :  |
Db        798 SAAAAATWAAAAAAAAAWATTTWTAAAAWATAATTTTTATABATTTWTTTTTWTWAAATTT 857

Qy        198 CTGAATCAATTAGGAATTTTTTTTTTT 225
                | |   :::  |||||||||||||:
Db        858 TAAATTTTTBYHMAAATTTTTTTTTTW 885


RESULT 12
BU722265/c
LOCUS       BU722265                542 bp    mRNA    linear   EST 23-OCT-2003
DEFINITION  SJMAFC03 SJM Schistosoma japonicum cDNA, mRNA sequence.
ACCESSION   BU722265
VERSION     BU722265.1  GI:28329618
KEYWORDS    EST.
SOURCE      Schistosoma japonicum
  ORGANISM  Schistosoma japonicum
            Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
            Strigeidida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE   1  (bases 1 to 542)
  AUTHORS   Hu,W., Yan,Q., Shen,D.K., Liu,F., Zhu,Z.D., Song,H.D., Xu,X.R.,
            Wang,Z.J., Rong,Y.P., Zeng,L.C., Wu,J., Zhang,X., Wang,J.J.,
            Xu,X.N., Wang,S.Y., Fu,G., Zhang,X.L., Wang,Z.Q., Brindley,P.J.,
            McManus,D.P., Xue,C.L., Feng,Z., Chen,Z. and Han,Z.G.
  TITLE     Evolutionary and biomedical implications of a Schistosoma japonicum
            complementary DNA resource
  JOURNAL   Nat. Genet. 35 (2), 139-147 (2003)
  MEDLINE   22879925
COMMENT     Contact: Zeguang Han
            Chinese National Human Genome Center at Shanghai
            351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai
            201203, P. R. China
            Tel: 86-21-50801919(ex.45)
            Fax: 86-21-50801922
            Email: hanzg@chgc.sh.cn.
FEATURES             Location/Qualifiers
     source          1. .542
                     /organism="Schistosoma japonicum"
                     /mol_type="mRNA"
                     /db_xref="taxon:6182"
                     /sex="male"
                     /tissue_type="Whole body"
                     /dev_stage="adult"
                     /lab_host="rabbits"
                     /clone_lib="SJM"
ORIGIN

  Query Match             18.8%;  Score 42.2;  DB 13;  Length 542;
  Best Local Similarity   59.2%;  Pred. No. 91;
  Matches   71;  Conservative    0;  Mismatches   49;  Indels    0;  Gaps    0;

Qy        104 TTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAACACCTTTTTAAG 163
              |||| |||||  | |    ||| ||| |||| | |||| ||||||  ||  ||
Db        542 TTTTAAAAGCCCCCCCCTCCCAAAATAACGGTAAAAAAAAAAAAACGAAATTTCCCTTA 483

Qy        164 AGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTT 223
              | ||  |    ||  |||| | |   |  |   |||||| ||  |  ||||||||||||
Db        482 AATTTTTACNAACCAATTTATTTTTGGTGGTGGCCTGAATTAAAAAAAAATTTTTTTTTT 423


RESULT 13
BX415058/c
LOCUS       BX415058                1056 bp    mRNA    linear   EST 15-MAY-2003
DEFINITION  BX415058 Homo sapiens THYMUS Homo sapiens cDNA clone CS0CAP004YG19
            3-PRIME, mRNA sequence.
ACCESSION   BX415058
VERSION     BX415058.1  GI:30767520
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 1056)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0CAP004AD10NP1.
FEATURES             Location/Qualifiers
     source          1. .1056
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0CAP004YG19"
                     /tissue_type="THYMUS"
                     /clone_lib="Homo sapiens THYMUS"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             18.8%;  Score 42.2;  DB 13;  Length 1056;
  Best Local Similarity   32.6%;  Pred. No. 58;
  Matches   70;  Conservative   48;  Mismatches   97;  Indels    0;  Gaps    0;

Qy          11 TTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTT 70
               :| |||    :|:::  |:: |::|  :||   :    |   |  | |  :  |    |||
Db         941 WTTTTTTTTNMAWWWTTTWWAAWWTTTWTTTAAWTTTTTTTTTTTTTTWTWYAAACAATTT 882

Qy          71 TACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAA 130
               |:  ||   || : ||:  | |||    |    :|::: : :  ::   : :     :::|
Db         881 TWWWTTATAAAAWTTAWAAAATTTTTTTTTWWWTWWWWWTWWWWWWWWWWWWWAWWWWWA 822

Qy         131 AACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCC 190
               ||  ||| |||| ||||||:::::::::::::| |    |    :: : :||: |
Db         821 AAAAAAAAAAAAAAAAAAAAHHHHHHHHWWWWAAAAAAAAAAAAWWWWWWTTWTANNNNNN 762

Qy         191 TCCTCTGCTGAATCAATTAGGAATTTTTTTTTTT 225
               | |  |  :  ||  |  : :||||||||||
Db         761 NATTTTTTTTTWAAAAAAAAAWTWTTTTTTTTTT 727


RESULT 14
BX462207/c
LOCUS       BX462207               1201 bp    mRNA    linear   EST 22-MAY-2003
DEFINITION  BX462207 Homo sapiens B CELLS (RAMOS CELL LINE) Homo sapiens cDNA
            clone CS0DG004YB03 5-PRIME, mRNA sequence.
ACCESSION   BX462207
VERSION     BX462207.1  GI:31023422
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1201)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Genoscope
            Genoscope - Centre National de Sequencage
            BP 191 91006 EVRY cedex - France
            Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
            Library was constructed by Life Technologies, a division of
            Invitrogen. This sequence belongs to sequence cluster 24.r For more
            information about this cluster, see http://www.genoscope.cns.fr/
            cgi-bin/cluster.cgi?seq=CS0DG004CA02QP1&cluster=24.r. Contact :
            Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue Genoscope sequence ID : CS0DG004CA02QP1.
FEATURES             Location/Qualifiers
     source          1. .1201
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DG004YB03"
                     /tissue_type="B CELLS (RAMOS CELL LINE)"
                     /cell_line="RAMOS CELL LINE"
                     /clone_lib="Homo sapiens B CELLS (RAMOS CELL LINE)"
                     /note="Vector: pCMVSPORT_6; 1st strand cDNA was primed
                     with a NotI-oligo(dT) primer. Five prime end enriched,
                     double-strand cDNA was digested with Not I and cloned into
                     the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                     Library was not normalized."
ORIGIN

  Query Match             18.7%;  Score 42;  DB 13;  Length 1201;
  Best Local Similarity   41.2%;  Pred. No. 58;
  Matches   84;  Conservative   25;  Mismatches   95;  Indels    0;  Gaps    0;

Qy          22 ATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAA 81
               :||||:|||  || |:|: |  :   :  |  | |   ||: ||  ||||   :: | |:
Db         729 WTAAAWTTTTTAAAAWAWAAAAWTWWWAAAAAAAAAAWTAWTTTTTTTTTTTTWWAAAAW 670

Qy          82 GAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAA 141
               || : ::: |:: |   |  |||:| |||   |   : :     ||||||   :| |||
Db         669 AAAAWAWWWTTWWTAAAAATTTTTWTWAAAAAAAAAAWTWAAAAWAAAAAAAAATWAAAAA 610

Qy         142 ACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGA 201
               | ||||||| ||       || |   ::    |  |:::  | |  |  | |  |
Db         609 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAWWAAAAAAAAAWWTTTTTTTTTTTTTTTTT 550

Qy         202 ATCAATTAGGAATTTTTTTTTTT 225
               |   ||     ||||||||||||
Db         549 TTTTTTTTTTTTTTTTTTTTTT 526


RESULT 15
CNS058MA/c
LOCUS       CNS058MA                964 bp    DNA     linear   GSS 01-SEP-2000
DEFINITION  Tetraodon nigroviridis genome survey sequence T3 end of clone
            002B22 of library A from Tetraodon nigroviridis, genomic survey
            sequence.
ACCESSION   AL326107
VERSION     AL326107.1  GI:8219696
KEYWORDS    GSS; genome survey sequence.
SOURCE      Tetraodon nigroviridis
  ORGANISM  Tetraodon nigroviridis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
            Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
            Tetradontoidea; Tetraodontidae; Tetraodon.
REFERENCE   1
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
            Bernot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
            Saurin,W. and Weissenbach,J.
  TITLE     Estimate of human gene number provided by genome-wide analysis
            using Tetraodon nigroviridis DNA sequence
  JOURNAL   Nat. Genet. 25 (2), 235-238 (2000)
  MEDLINE   20296633
   PUBMED   10835645
REFERENCE   2
  AUTHORS   Roest Crollius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C.,
            Fizames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F.,
            Saurin,W., Bernot,A. and Weissenbach,J.
  TITLE     Characterization and repeat analysis of the compact genome of the
            freshwater pufferfish Tetraodon nigroviridis
  JOURNAL   Genome Res. 10 (7), 939-949 (2000)
  MEDLINE   20359837
   PUBMED   10899143
REFERENCE   3  (bases 1 to 964)
```

```
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-APR-2000) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     This sequence is a single read and was generated as part of a large
            scale clone-end sequencing project of the Tetraodon nigroviridis
            genome. For more information, please take a look at
            http://www.genoscope.cns.fr/Tetraodon.
FEATURES             Location/Qualifiers
     source          1. .964
                     /organism="Tetraodon nigroviridis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:99883"
                     /clone="002B22"
                     /clone_lib="A"
                     /note="Genoscope sequence ID : C0AA002DA11A1~end : T3"
ORIGIN

  Query Match             18.6%;  Score 41.8;  DB 29;  Length 964;
  Best Local Similarity   52.4%;  Pred. No. 74;
  Matches   88;  Conservative    0;  Mismatches   80;  Indels    0;  Gaps    0;

Qy          4 TGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATA 63
              |||| | ||       |      | ||| |      |  || | |     |  |||| || |
Db        821 TGGTGATTTNCCGGTTGTGACCATTTTTAGCCGTACAATTATGGGATTTTCTATGTTAAA 762

Qy         64 TTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCC 123
                || ||| || |  |  || |||| | |||   | | || || ||  |  | |||
Db        761 NATTTTCTACATTAAACATTATGTTTAATGTCATGAATGATTGTCNAAAATTAGTTGTAT 702

Qy        124 CCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATG 171
                ||||||  |||||||| |  || ||||  | |||   || |  ||
Db        701 TTAAAAAAAAAAAACAAAAAATCAAGACAAGTTCTTTTTCAGATTCTG 654


Search completed: July 27, 2004, 23:03:09
Job time : 2781 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 17:03:10 ; Search time 3655 Seconds
(without alignments)
2668.176 Million cell updates/sec

Title: US-09-765-231A-58
Perfect score: 225
Sequence: 1 tgatggtaagttgtttcagg.....attaggaatttttttttttt 225

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	225	100.0	225	6	AX202128	AX202128 Sequence
c 2	225	100.0	143800	2	AC011875	AC011875 Homo sapi
c 3	225	100.0	186965	9	AP000848	AP000848 Homo sapi
4	175.6	78.0	172830	2	AP001320	AP001320 Homo sapi
5	128.4	57.1	75002	2	AC023384	AC023384 Homo sapi
c 6	47.2	21.0	85916	3	AC117080	AC117080 Dictyoste
7	46.2	20.5	252248	2	AC094553	AC094553 Rattus no
8	45.8	20.4	110000	2	PFMAL13_04	Continuation (5 of
c 9	45.2	20.1	9971	1	U67577	U67577 Methanococc
10	45.2	20.1	110000	6	AR271569_01	Continuation (2 of
11	45.2	20.1	231461	2	AC096278	AC096278 Rattus no
c 12	45	20.0	170419	9	AC146265	AC146265 Pan trogl
c 13	44.4	19.7	147727	10	AL929001	AL929001 Mouse DNA
c 14	44	19.6	148750	2	AC104893	AC104893 Mus muscu
c 15	43.8	19.5	200110	10	AC117237	AC117237 Mus muscu
c 16	43	19.1	182163	2	BX000690	BX000690 Danio rer
c 17	42.6	18.9	127902	8	AP005406	AP005406 Oryza sat
c 18	42.4	18.8	199551	2	AC006281	AC006281 Plasmodiu
c 19	42.4	18.8	251551	3	AE014844	AE014844 Plasmodiu
20	42	18.7	147760	9	AC011846	AC011846 Homo sapi
c 21	42	18.7	151802	3	AC114263	AC114263 Dictyoste
c 22	41.8	18.6	156140	9	AC067745	AC067745 Homo sapi
23	41.8	18.6	215467	2	AC013420	AC013420 Homo sapi
c 24	41.8	18.6	215734	2	AC073828	AC073828 Mus muscu
25	41.6	18.5	171050	9	AC112232	AC112232 Homo sapi
26	41.6	18.5	181864	2	AC027460	AC027460 Homo sapi
c 27	41.6	18.5	194127	2	BX571853	BX571853 Danio rer
28	41.6	18.5	257650	2	BX323874	BX323874 Danio rer
29	41.4	18.4	110000	2	PFMAL7P1_03	Continuation (4 of
c 30	41.2	18.3	138564	10	AL645950	AL645950 Mouse DNA
c 31	41	18.2	5518	6	AX323692	AX323692 Sequence
32	41	18.2	169546	2	AC004157	AC004157 Plasmodiu
c 33	40.8	18.1	130355	8	AP003412	AP003412 Oryza sat
34	40.8	18.1	190561	2	AC118246	AC118246 Mus muscu
c 35	40.8	18.1	192752	2	AC118028	AC118028 Mus muscu
c 36	40.8	18.1	228071	10	AC115750	AC115750 Mus muscu
c 37	40.6	18.0	158764	10	AC132083	AC132083 Mus muscu
c 38	40.6	18.0	160429	2	AC136950	AC136950 Homo sapi
39	40.6	18.0	203241	9	AC008282	AC008282 Homo sapi
40	40.4	18.0	12300	10	MMU84903	U84903 Mus musculu
c 41	40.4	18.0	101491	10	AP003183	AP003183 Mus muscu
c 42	40.4	18.0	158123	10	AC134832	AC134832 Mus muscu
43	40.4	18.0	197307	2	AC130714	AC130714 Mus muscu
c 44	40.4	18.0	198600	2	AC133496	AC133496 Mus muscu
45	40.4	18.0	199371	2	AC147220	AC147220 Mus muscu

ALIGNMENTS

RESULT 1
AX202128
LOCUS AX202128 225 bp DNA linear PAT 30-AUG-2001
DEFINITION Sequence 58 from Patent WO0153531.
ACCESSION AX202128
VERSION AX202128.1 GI:15391919
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Phippard,D., Vasanthakamur,G., Dotson,S. and Ma,X.J.
TITLE Osteoarthritis tissue derived nucleic acids, polypeptides, vectors, and cells

```
  JOURNAL    Patent: WO 0153531-A 58 26-JUL-2001;
             Pharmacia Corporation (US)
FEATURES               Location/Qualifiers
     source            1. .225
                       /organism="Homo sapiens"
                       /mol_type="unassigned DNA"
                       /db_xref="taxon:9606"
ORIGIN

  Query Match             100.0%;  Score 225;  DB 6;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 1.8e-40;
  Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60

Qy         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120

Qy        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180

Qy        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225


RESULT 2
AC011875/c
LOCUS       AC011875              143800 bp    DNA     linear   HTG 12-MAR-2000
DEFINITION  Homo sapiens clone RP11-16K5, WORKING DRAFT SEQUENCE, 26 unordered
            pieces.
ACCESSION   AC011875
VERSION     AC011875.3  GI:7107950
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143800)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-16K5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143800)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-1999) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 28, 2000 this sequence version replaced gi:6453961.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu

            -------------- Project Information
                Center project name: L3566
                Center clone name: 16_K_5
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 93387 bases at least Q40
                Consensus quality: 115701 bases at least Q30
                Consensus quality: 130381 bases at least Q20
                Insert size: 141000; agarose-fp
                Insert size: 141300; sum-of-contigs
                Quality coverage: 3.7 in Q20 bases; agarose-fp
                Quality coverage: 3.7 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1029: contig of 1029 bp in length
            *     1030     1129: gap of 100 bp
            *     1130     2964: contig of 1835 bp in length
            *     2965     3064: gap of 100 bp
            *     3065     4253: contig of 1189 bp in length
            *     4254     4353: gap of 100 bp
            *     4354     5678: contig of 1325 bp in length
            *     5679     5778: gap of 100 bp
            *     5779     6855: contig of 1077 bp in length
            *     6856     6955: gap of 100 bp
            *     6956     8847: contig of 1892 bp in length
            *     8848     8947: gap of 100 bp
            *     8948    10945: contig of 1998 bp in length
            *    10946    11045: gap of 100 bp
            *    11046    13001: contig of 1956 bp in length
            *    13002    13101: gap of 100 bp
            *    13102    15430: contig of 2329 bp in length
            *    15431    15530: gap of 100 bp
            *    15531    17919: contig of 2389 bp in length
            *    17920    18019: gap of 100 bp
            *    18020    20135: contig of 2116 bp in length
            *    20136    20235: gap of 100 bp
            *    20236    23750: contig of 3515 bp in length
            *    23751    23850: gap of 100 bp
            *    23851    26794: contig of 2944 bp in length
            *    26795    26894: gap of 100 bp
            *    26895    29482: contig of 2588 bp in length
            *    29483    29582: gap of 100 bp
            *    29583    33174: contig of 3592 bp in length
            *    33175    33274: gap of 100 bp
            *    33275    35185: contig of 1911 bp in length
            *    35186    35285: gap of 100 bp
            *    35286    39745: contig of 4460 bp in length
            *    39746    39845: gap of 100 bp
            *    39846    44222: contig of 4377 bp in length
            *    44223    44322: gap of 100 bp
            *    44323    48990: contig of 4668 bp in length
            *    48991    49090: gap of 100 bp
            *    49091    57790: contig of 8700 bp in length
            *    57791    57890: gap of 100 bp
            *    57891    66822: contig of 8932 bp in length
            *    66823    66922: gap of 100 bp
            *    66923    76709: contig of 9787 bp in length
            *    76710    76809: gap of 100 bp
            *    76810    92865: contig of 16056 bp in length
            *    92866    92965: gap of 100 bp
            *    92966   106278: contig of 13313 bp in length
            *   106279   106378: gap of 100 bp
            *   106379   123041: contig of 16663 bp in length
            *   123042   123141: gap of 100 bp
```

```
          *   123142   143800: contig of 20659 bp in length.
FEATURES             Location/Qualifiers
     source          1. .143800
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-16K5"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1029
                     /note="assembly_fragment"
     misc_feature    1130. .2964
                     /note="assembly_fragment"
     misc_feature    3065. .4253
                     /note="assembly_fragment"
     misc_feature    4354. .5678
                     /note="assembly_fragment"
     misc_feature    5779. .6855
                     /note="assembly_fragment"
     misc_feature    6956. .8847
                     /note="assembly_fragment"
     misc_feature    8948. .10945
                     /note="assembly_fragment"
     misc_feature    11046. .13001
                     /note="assembly_fragment"
     misc_feature    13102. .15430
                     /note="assembly_fragment"
     misc_feature    15531. .17919
                     /note="assembly_fragment"
     misc_feature    18020. .20135
                     /note="assembly_fragment"
     misc_feature    20236. .23750
                     /note="assembly_fragment"
     misc_feature    23851. .26794
                     /note="assembly_fragment"
     misc_feature    26895. .29482
                     /note="assembly_fragment"
     misc_feature    29583. .33174
                     /note="assembly_fragment"
     misc_feature    33275. .35185
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    35286. .39745
                     /note="assembly_fragment"
     misc_feature    39846. .44222
                     /note="assembly_fragment"
     misc_feature    44323. .48990
                     /note="assembly_fragment"
     misc_feature    49091. .57790
                     /note="assembly_fragment"
     misc_feature    57891. .66822
                     /note="assembly_fragment"
     misc_feature    66923. .76709
                     /note="assembly_fragment"
     misc_feature    76810. .92865
                     /note="assembly_fragment"
     misc_feature    92966. .106278
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    106379. .123041
                     /note="assembly_fragment"
     misc_feature    123142. .143800
                     /note="assembly_fragment"
ORIGIN

  Query Match             100.0%;  Score 225;  DB 2;  Length 143800;
  Best Local Similarity   100.0%;  Pred. No. 6e-41;
  Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50974 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 50915

Qy         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50914 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 50855

Qy        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      50854 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 50795

Qy        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db      50794 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTT 50750


RESULT 3
AP000848/c
LOCUS       AP000848               186965 bp    DNA     linear   PRI 17-JUN-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-775D16, complete
            sequence.
ACCESSION   AP000848
VERSION     AP000848.5  GI:31790708
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (1999)
REFERENCE   2  (bases 1 to 186965)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jun 16, 2003 this sequence version replaced gi:14517575.
FEATURES             Location/Qualifiers
     source          1. .186965
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-775D16"
ORIGIN

  Query Match             100.0%;  Score 225;  DB 9;  Length 186965;
  Best Local Similarity   100.0%;  Pred. No. 5.7e-41;
  Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83181 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 83122

Qy         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83121 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 83062

Qy        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      83061 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 83002

Qy        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db      83001 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTT 82957


RESULT 4
```

```
AP001320
LOCUS       AP001320              172830 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-799O4 map 11q14, WORKING
            DRAFT SEQUENCE, 32 unordered pieces.
ACCESSION   AP001320
VERSION     AP001320.2  GI:8117247
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 172830)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 172,830 genomic DNA of 11q14
  JOURNAL   Published Only in DataBase (2000)
REFERENCE   2  (bases 1 to 172830)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 30, 2000 this sequence version replaced gi:7209763.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: RP11-799O4
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 145836 bases at least Q40
                Consensus quality: 159404 bases at least Q30
                Consensus quality: 166388 bases at least Q20
                Insert size: 169730; sum-of-contigs
                Quality coverage: 4.17x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            32 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                  1     15732 contig of  15732 bp in length
              15833     34101 contig of  18269 bp in length
              34202     47479 contig of  13278 bp in length
              47580     57756 contig of  10177 bp in length
              57857     69426 contig of  11570 bp in length
              69527     80185 contig of  10659 bp in length
              80286     88989 contig of   8704 bp in length
              89090     99362 contig of  10273 bp in length
              99463    103433 contig of   3971 bp in length
             103534    109064 contig of   5531 bp in length
             109165    113855 contig of   4691 bp in length
             113956    119140 contig of   5185 bp in length
             119241    124661 contig of   5421 bp in length
             124762    126822 contig of   2061 bp in length
             126923    131337 contig of   4415 bp in length
             131438    135078 contig of   3641 bp in length
             135179    138134 contig of   2956 bp in length
             138235    142123 contig of   3889 bp in length
             142224    144719 contig of   2496 bp in length
             144820    149030 contig of   4211 bp in length
             149131    151134 contig of   2004 bp in length
             151235    153751 contig of   2517 bp in length
             153852    156504 contig of   2653 bp in length
             156605    159583 contig of   2979 bp in length
             159684    161669 contig of   1986 bp in length
             161770    164135 contig of   2366 bp in length
             164236    165854 contig of   1619 bp in length
             165955    167495 contig of   1541 bp in length
             167596    168771 contig of   1176 bp in length
             168872    170322 contig of   1451 bp in length
             170423    171562 contig of   1140 bp in length
             171663    172830 contig of   1168 bp in length
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 32 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*        1     15732: contig of 15732 bp in length
*    15733     15832: gap of 100 bp
*    15833     34101: contig of 18269 bp in length
*    34102     34201: gap of 100 bp
*    34202     47479: contig of 13278 bp in length
*    47480     47579: gap of 100 bp
*    47580     57756: contig of 10177 bp in length
*    57757     57856: gap of 100 bp
*    57857     69426: contig of 11570 bp in length
*    69427     69526: gap of 100 bp
*    69527     80185: contig of 10659 bp in length
*    80186     80285: gap of 100 bp
*    80286     88989: contig of 8704 bp in length
*    88990     89089: gap of 100 bp
*    89090     99362: contig of 10273 bp in length
*    99363     99462: gap of 100 bp
*    99463    103433: contig of 3971 bp in length
*   103434    103533: gap of 100 bp
*   103534    109064: contig of 5531 bp in length
*   109065    109164: gap of 100 bp
*   109165    113855: contig of 4691 bp in length
*   113856    113955: gap of 100 bp
*   113956    119140: contig of 5185 bp in length
*   119141    119240: gap of 100 bp
*   119241    124661: contig of 5421 bp in length
*   124662    124761: gap of 100 bp
*   124762    126822: contig of 2061 bp in length
*   126823    126922: gap of 100 bp
*   126923    131337: contig of 4415 bp in length
*   131338    131437: gap of 100 bp
*   131438    135078: contig of 3641 bp in length
*   135079    135178: gap of 100 bp
*   135179    138134: contig of 2956 bp in length
*   138135    138234: gap of 100 bp
*   138235    142123: contig of 3889 bp in length
*   142124    142223: gap of 100 bp
*   142224    144719: contig of 2496 bp in length
*   144720    144819: gap of 100 bp
*   144820    149030: contig of 4211 bp in length
*   149031    149130: gap of 100 bp
*   149131    151134: contig of 2004 bp in length
*   151135    151234: gap of 100 bp
*   151235    153751: contig of 2517 bp in length
*   153752    153851: gap of 100 bp
*   153852    156504: contig of 2653 bp in length
*   156505    156604: gap of 100 bp
*   156605    159583: contig of 2979 bp in length
*   159584    159683: gap of 100 bp
*   159684    161669: contig of 1986 bp in length
*   161670    161769: gap of 100 bp
*   161770    164135: contig of 2366 bp in length
*   164136    164235: gap of 100 bp
*   164236    165854: contig of 1619 bp in length
```

```
            *   165855   165954: gap of 100 bp
            *   165955   167495: contig of 1541 bp in length
            *   167496   167595: gap of 100 bp
            *   167596   168771: contig of 1176 bp in length
            *   168772   168871: gap of 100 bp
            *   168872   170322: contig of 1451 bp in length
            *   170323   170422: gap of 100 bp
            *   170423   171562: contig of 1140 bp in length
            *   171563   171662: gap of 100 bp
            *   171663   172830: contig of 1168 bp in length.
FEATURES             Location/Qualifiers
     source          1. .172830
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q14"
                     /clone="RP11-799O4"
     misc_feature    1. .15732
                     /note="assembly_fragment"
     misc_feature    15833. .34101
                     /note="assembly_fragment"
     misc_feature    34202. .47479
                     /note="assembly_fragment"
     misc_feature    47580. .57756
                     /note="assembly_fragment"
     misc_feature    57857. .69426
                     /note="assembly_fragment"
     misc_feature    69527. .80185
                     /note="assembly_fragment"
     misc_feature    80286. .88989
                     /note="assembly_fragment"
     misc_feature    89090. .99362
                     /note="assembly_fragment"
     misc_feature    99463. .103433
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    103534. .109064
                     /note="assembly_fragment"
     misc_feature    109165. .113855
                     /note="assembly_fragment"
     misc_feature    113956. .119140
                     /note="assembly_fragment"
     misc_feature    119241. .124661
                     /note="assembly_fragment"
     misc_feature    124762. .126822
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    126923. .131337
                     /note="assembly_fragment"
     misc_feature    131438. .135078
                     /note="assembly_fragment"
     misc_feature    135179. .138134
                     /note="assembly_fragment"
     misc_feature    138235. .142123
                     /note="assembly_fragment"
     misc_feature    142224. .144719

  Query Match             78.0%;  Score 175.6;  DB 2;  Length 172830;
  Best Local Similarity   88.9%;  Pred. No. 6.7e-30;
  Matches  201;  Conservative    0;  Mismatches   24;  Indels    1;  Gaps    1;

Qy          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      28909 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 28968

Qy         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCT-ACGCT 119
              ||||| | ||||||||||||||||||||||||||||||||||||||||||| || ||
Db      28969 ATATTTGGTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTAACCCT 29028

Qy        120 GTCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCA 179
              ||||||| ||| ||| ||| ||||   ||||| ||| ||||||||||||||||||| |||
Db      29029 GTCCCCCCAAATACGGAACCAAACCCTAAAACCACCCTTTTAAGAGTTGATGGCTAATCA 29088

Qy        180 TTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||| |||||||| ||||| |||| |  |   ||||| |||||
Db      29089 TTTGATTTGCCTCCTGTGCTGGATCACATTAGGGATTTTTCTTTTT 29134
```

RESULT 5
AC023384

```
LOCUS       AC023384               75002 bp    DNA     linear   HTG 13-JUL-2000
DEFINITION  Homo sapiens chromosome 11 clone RP11-589I12 map 11, LOW-PASS
            SEQUENCE SAMPLING.
ACCESSION   AC023384
VERSION     AC023384.2  GI:9144035
KEYWORDS    HTG; HTGS_PHASE0.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 75002)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-589I12
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 75002)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
            Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
            McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
            Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
            Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
            Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
            Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
            Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
            Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 13, 2000 this sequence version replaced gi:6970532.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L5114
                Center clone name: 589_I_12
            --------------.
            * NOTE: This record contains 88 individual
            * sequencing reads that have not been assembled into
            * contigs. Runs of N are used to separate the reads
            * and the order in which they appear is completely
            * arbitrary. Low-pass sequence sampling is useful for
            * identifying clones that may be gene-rich and allows
            * overlap relationships among clones to be deduced.
            * However, it should not be assumed that this clone
            * will be sequenced to completion. In the event that
            * the record is updated, the accession number will
            * be preserved.
            *        1      724: contig of 724 bp in length
            *      725      824: gap of 100 bp
            *      825     1573: contig of 749 bp in length
            *     1574     1673: gap of 100 bp
```

```
*     1674     2455: contig of 782 bp in length
*     2456     2555: gap of 100 bp
*     2556     3328: contig of 773 bp in length
*     3329     3428: gap of 100 bp
*     3429     4179: contig of 751 bp in length
*     4180     4279: gap of 100 bp
*     4280     5022: contig of 743 bp in length
*     5023     5122: gap of 100 bp
*     5123     5865: contig of 743 bp in length
*     5866     5965: gap of 100 bp
*     5966     6703: contig of 738 bp in length
*     6704     6803: gap of 100 bp
*     6804     7554: contig of 751 bp in length
*     7555     7654: gap of 100 bp
*     7655     8408: contig of 754 bp in length
*     8409     8508: gap of 100 bp
*     8509     9254: contig of 746 bp in length
*     9255     9354: gap of 100 bp
*     9355    10118: contig of 764 bp in length
*    10119    10218: gap of 100 bp
*    10219    10978: contig of 760 bp in length
*    10979    11078: gap of 100 bp
*    11079    11850: contig of 772 bp in length
*    11851    11950: gap of 100 bp
*    11951    12712: contig of 762 bp in length
*    12713    12812: gap of 100 bp
*    12813    13553: contig of 741 bp in length
*    13554    13653: gap of 100 bp
*    13654    14392: contig of 739 bp in length
*    14393    14492: gap of 100 bp
*    14493    15244: contig of 752 bp in length
*    15245    15344: gap of 100 bp
*    15345    16095: contig of 751 bp in length
*    16096    16195: gap of 100 bp
*    16196    16927: contig of 732 bp in length
*    16928    17027: gap of 100 bp
*    17028    17766: contig of 739 bp in length
*    17767    17866: gap of 100 bp
*    17867    18615: contig of 749 bp in length
*    18616    18715: gap of 100 bp
*    18716    19489: contig of 774 bp in length
*    19490    19589: gap of 100 bp
*    19590    20344: contig of 755 bp in length
*    20345    20444: gap of 100 bp
*    20445    21205: contig of 761 bp in length
*    21206    21305: gap of 100 bp
*    21306    22089: contig of 784 bp in length
*    22090    22189: gap of 100 bp
*    22190    22972: contig of 783 bp in length
*    22973    23072: gap of 100 bp
*    23073    23826: contig of 754 bp in length
*    23827    23926: gap of 100 bp
*    23927    24674: contig of 748 bp in length
*    24675    24774: gap of 100 bp
*    24775    25507: contig of 733 bp in length
*    25508    25607: gap of 100 bp
*    25608    26347: contig of 740 bp in length
*    26348    26447: gap of 100 bp
*    26448    27202: contig of 755 bp in length
*    27203    27302: gap of 100 bp
*    27303    28052: contig of 750 bp in length
*    28053    28152: gap of 100 bp
*    28153    28909: contig of 757 bp in length
*    28910    29009: gap of 100 bp
*    29010    29741: contig of 732 bp in length
*    29742    29841: gap of 100 bp
*    29842    30624: contig of 783 bp in length
*    30625    30724: gap of 100 bp
*    30725    31497: contig of 773 bp in length
*    31498    31597: gap of 100 bp
*    31598    32373: contig of 776 bp in length
*    32374    32473: gap of 100 bp
*    32474    33222: contig of 749 bp in length
*    33223    33322: gap of 100 bp
*    33323    34072: contig of 750 bp in length
*    34073    34172: gap of 100 bp
*    34173    34932: contig of 760 bp in length
*    34933    35032: gap of 100 bp
*    35033    35791: contig of 759 bp in length
*    35792    35891: gap of 100 bp
*    35892    36636: contig of 745 bp in length
*    36637    36736: gap of 100 bp
*    36737    37468: contig of 732 bp in length
*    37469    37568: gap of 100 bp
*    37569    38328: contig of 760 bp in length
*    38329    38428: gap of 100 bp
*    38429    39191: contig of 763 bp in length
*    39192    39291: gap of 100 bp
*    39292    40036: contig of 745 bp in length
*    40037    40136: gap of 100 bp
*    40137    40895: contig of 759 bp in length
*    40896    40995: gap of 100 bp
*    40996    41729: contig of 734 bp in length
*    41730    41829: gap of 100 bp
*    41830    42577: contig of 748 bp in length
*    42578    42677: gap of 100 bp
*    42678    43414: contig of 737 bp in length
*    43415    43514: gap of 100 bp
*    43515    44267: contig of 753 bp in length
*    44268    44367: gap of 100 bp
*    44368    45123: contig of 756 bp in length
*    45124    45223: gap of 100 bp
*    45224    45973: contig of 750 bp in length
*    45974    46073: gap of 100 bp
*    46074    46816: contig of 743 bp in length
*    46817    46916: gap of 100 bp
*    46917    47681: contig of 765 bp in length
*    47682    47781: gap of 100 bp
*    47782    48544: contig of 763 bp in length
*    48545    48644: gap of 100 bp
*    48645    49410: contig of 766 bp in length
*    49411    49510: gap of 100 bp
*    49511    50247: contig of 737 bp in length
*    50248    50347: gap of 100 bp
*    50348    51067: contig of 720 bp in length
*    51068    51167: gap of 100 bp
*    51168    51915: contig of 748 bp in length
*    51916    52015: gap of 100 bp
*    52016    52745: contig of 730 bp in length
*    52746    52845: gap of 100 bp
*    52846    53584: contig of 739 bp in length
*    53585    53684: gap of 100 bp
*    53685    54427: contig of 743 bp in length
*    54428    54527: gap of 100 bp
*    54528    55292: contig of 765 bp in length
*    55293    55392: gap of 100 bp
*    55393    56164: contig of 772 bp in length
*    56165    56264: gap of 100 bp
*    56265    56985: contig of 721 bp in length
*    56986    57085: gap of 100 bp
*    57086    57829: contig of 744 bp in length
*    57830    57929: gap of 100 bp
*    57930    58677: contig of 748 bp in length

  Query Match             57.1%;  Score 128.4;  DB 2;  Length 75002;
  Best Local Similarity   98.6%;  Pred. No. 2.9e-19;
  Matches  140;  Conservative    0;  Mismatches    1;  Indels    1;  Gaps    1;

Qy         84 ATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAAC 143
              ||||||||||||||||||||||||||||||||||||||| ||| ||||||||||||
Db      35932 ATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCC-AAAGACGAAACAAAAC 35990

Qy        144 AAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAAT 203
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      35991 AAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAAT 36050
```

```
Qy        204 CAATTAGGAATTTTTTTTTTT 225
              |||||||||||||||||||||
Db      36051 CAATTAGGAATTTTTTTTTTT 36072


RESULT 6
AC117080/c
LOCUS       AC117080               85916 bp    DNA     linear   INV 12-MAR-2003
DEFINITION  Dictyostelium discoideum chromosome 2 map complement(821514-735598)
            strain AX4, complete sequence.
ACCESSION   AC117080 AC114260
VERSION     AC117080.2  GI:28850277
KEYWORDS    HTG.
SOURCE      Dictyostelium discoideum
  ORGANISM  Dictyostelium discoideum
            Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
REFERENCE   1  (bases 1 to 85916)
  AUTHORS   Gloeckner,G., Eichinger,L., Szafranski,K., Pachebat,J., Dear,P.,
            Lehmann,R., Baumgart,C., Parra,G., April,J.F., Guigo,R., Kumpf,K.,
            Tunggal,B., Cox,E., Quail,M.A., Platzer,M., Rosenthal,A. and
            Noegel,A.A.
  TITLE     Sequence and analysis of chromosome 2 of Dictyostelium discoideum
  JOURNAL   Nature 418 (6893), 79-85 (2002)
  MEDLINE   22092622
   PUBMED   12097910
  REMARK    The Dictyostelium Genome Sequencing Consortium
REFERENCE   2  (bases 1 to 85916)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-APR-2002) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   3  (bases 1 to 85916)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
REFERENCE   4  (bases 1 to 85916)
  AUTHORS   Baumgart,C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-MAR-2003) Genome Analysis, Institute of Molecular
            Biotechnology, Beutenberstr. 11, Jena 07745, Germany
COMMENT     On or before Mar 5, 2003 this sequence version replaced
            gi:19424360, gi:20066270.
            CDS predictions from GeneID do not necessarily reflect true genes.
            Further Information is available from IMB Jena, Department of
            Genome Analysis
            (http://genome.imb-jena.de/dictyostelium/)
            and the Univerity Colonge, Institute for Biochemistry I
            (http://www.uni-koeln.de/dictyostelium/project.shtml
            Funding
            Agency : Deutsche Forschungsgemeinschaft (DFG).
FEATURES              Location/Qualifiers
     source           1. .85916
                      /organism="Dictyostelium discoideum"
                      /mol_type="genomic DNA"
                      /strain="AX4"
                      /db_xref="taxon:44689"
                      /chromosome="2"
                      /map="complement(821514-735598)"
     CDS              join(9. .188,279. .1337)
                      /note="GeneID exon scores (in order of location ranges):
                      9.57, 89.77 - GSCJ_ID dd_00785"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO53082.1"
                      /db_xref="GI:28850278"
                      /translation="MEDNNNSSSNENKQNELSTPSIKKERHLVIMQHGLHGTSLDFKT
                      IRNHFLKQKHLDNCIFISANSNSHFLATHDGIDKIGERLFNEVKELYEQYDHPEKISM
                      IGHSLGGLITRYAIGLLYDDGFFKKCKPDQFISLSSPHCGSRRPSTTIFNKVAHIFVD
                      NFLSVTGKQLILHDTEIPDNIKTFPSTSSPPPNEKLKSSKTIVNSSVKNETDLSLPLA
                      EAKEPSIYKEVGNNEKLMIIEKKEENEIITNDQEVPMPLLVRLSEGIFFNGLNSFRKR
                      TLYSNIYNDVQVNFCTSDISAKNPYTLGKLMKFSEKYRHIIEEESILDIDPNLLEQQS
                      PPPDKKPFDEKDLDEYFTHDTHHHFLKRILKNLNQLHFVRYHMYFKNMLSHTNIIVKR
                      EWINSEGFEIIEHLVSHFEG"
     CDS              join(2619. .2689,2806. .2917,3021. .3235,3389. .3938)
                      /note="GeneID exon scores (in order of location ranges):
                      4.00, 11.19, 30.82, 62.55 - GSCJ_ID dd_00787"
                      /codon_start=1
                      /product="similar to Mus musculus (Mouse). 26S proteasome
                      regulatory subunit S5A (Rpn10) (Multiubiquitin chain
                      binding protein)"
                      /protein_id="AAO53083.1"
                      /db_xref="GI:28850279"
                      /translation="MYTASKTQSNPESAVSIMSMAGKKPEVLVTLTQELSKILSGAQE
                      LKINGKIDFSTTMQIAQLALRHRQNNHQHPRIIAFVGSPLKETKEELIQLAKRLKKNG
                      VAVDIINFGEVTENSDKLEAFINDVNNNDESHLLTVPPGPHILSDIILQSPIVESGSG
                      QFGSEFINADTDPDLAMALKLSLEEEKQRQERERKAREEANGGSTNSGTTTTTAPTES
                      NMDVNFEDDPELAEALALSMATDKMEVQSSTTNTDSQPPQQQQQPPTDDTSSEAFKDQ
                      DFLNSTLNSLPGVDPNRIKNALENLSKKDEDKDKDNEKK"
     CDS              complement(join(4186. .5395,5501. .5689,5854. .6303,
                      6445. .7055))
                      /note="GeneID exon scores (in order of location ranges):
                      94.07, 12.83, 33.44, 41.83 - GSCJ_ID dd_00788"
                      /codon_start=1
                      /product="similar to Mus musculus (Mouse). 12 days embryo
                      male wolffian duct includes surrounding region cDNA, RIKEN
                      full-length enriched library, clone:6720456H20
                      product:similar to  KAIA4653 PROTEIN (Fragment)"
                      /protein_id="AAM45299.2"
                      /db_xref="GI:28850280"
                      /translation="MNNSPTLSNTRNRKNNNNNSNSNSNSNNNNNNNNNNNNSNNNN
                      NNNNNVNRNVNNRNNNNNNIINVNNNVSSTTMNMAKDFMNWNDGIIEREDTQLQQQKD
                      RCFKILKKSTIACIVLLFISCTIIYSMTQYPSVSGGDAGELIVVAHQFGVAHPPGYPL
                      FTFLGYIFSHIIPSNTLSVAWKISFMSSMIGSIASIFIYLTVYLWVNNHWCGLLSAYM
                      FTFSPLIWMYQIQGEVFSMNNMFVAMLMFLGVWYTRVRIFENERYNTAFWTSERIAYL
                      SAFSCAIGLTNQHTLVLIVIPFAFWLMFISGRDQLWNIKILSNLVFYTLIGLSPYLLL
                      FITPKLNRVKYSWGNTSTLRGFIKHFLREEYGTLQLYGGDGGVISLFTKISIYIENLI
                      IQFGYIGLALSLIGLLNLLLGYNIRTFKWKSFGTMIIFSFLFYITFFFNLCNLPIDKP
                      LYRGVFLRFFMQPNVIISITMGLGIKSIFGFLNRIQRDGGTAATTTTTISKIQKYLLP
                      IIIILLVGNQIGFNYNLQDQSDNYSFYDYGHSVLDGLPRNTLLLVGGDLVTNVPMYLH
                      LCEKVRPDIDILSMEIMSWEWFKITQSPLFQRVKFPGNVYHPYIPDGYSLKDFLDANI
                      NDRPIYIGGDFKSGDSSFQNDYYTISKGLTSQIIPKKDSHKFNTFKIIKTTFSTFPSF
                      HIPNNTEKYPNDSWEHFMMEEMAVSLERAAETLLKEYLSQQNTESFKALELSVEILEK
                      ALTYVNDKCWSLKHLGICFDHLRYRVVQNVNNNNQQAQNKASNYSKQLLYFWKKYINQ
                      CTHEENTDQDWETIKKVIQLI"
     CDS              7833. .8195
                      /note="GeneID exon scores (in order of location ranges):
                      30.55 - GSCJ_ID dd_00789"
                      /codon_start=1
                      /product="hypothetical protein"
                      /protein_id="AAO53084.1"
                      /db_xref="GI:28850281"
                      /translation="MINPLNKPVLGTRLMNSVFFPNIKLQVIRNSGKTFENTNKIMFK
                      TDCSVGKTDIKTYMKALYDVNVDKVNTINVQGRIKSKTKGLQKRSKLNSKYKTPDYKK
                      AIITVDPSLRAQLSRGKN"
     CDS              9000. .9674
                      /note="GeneID exon scores (in order of location ranges):
                      54.39 - GSCJ_ID dd_02815"
                      /codon_start=1
                      /product="similar to Homo sapiens (Human). Small GTP
                      binding protein RAB32"
                      /protein_id="AAM45300.2"
                      /db_xref="GI:28850282"
                      /translation="MYSNKNDKDKDKDQNNENNKNNDDEAISLKILVVGKLACGKTSI
                      IQRYCHNNFQPKYKPTIGVDFQQKVLEIMGQKVLLQLWDIAGQERFGHMTRVYFQNAH
                      GAVIVFDATRSGTFLGAKAWKDDIDYCFNNENLPTILLANKCDLLTPPYTFPEDINTF
                      CEQNRFNKYFYTSAKEDTGINEALEELVKIILESYQTQEQSTGFKLSDQSQSTETTPT
                      QSKTCC"
     CDS              complement(10728. .12734)
                      /note="GeneID exon scores (in order of location ranges):
                      120.50 - GSCJ_ID dd_03061"
                      /codon_start=1
                      /product="similar to Plasmodium falciparum. Hypothetical
                      protein"
                      /protein_id="AAO53085.1"
                      /db_xref="GI:28850283"
```

/translation="MNNGKILVLNSLPNYLQKRIIQLLCEHIDFNNKQLVVNCNFPKI
LEYRENESLIKRLIINISLVSHQWFQIISNYLNVSVDFNFNSINNSYYSIIKATNVKI
LKIHYNHQNNYFYKFIKGNEFKLTINEINEKIKQLPNLKQLLIRNINISKVINDIDGT
IDTSNNSSYSNINIEGNFKIETNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNGDGDGDGDGNSDTTTNSITNINNFNNNSDNYHYNNRIINNDNLNFLKDIEFKGN
IELLEVEINKSNSMERLMELKSKISKVFSLNVVNTLSKELLFETFKNWLPSIRQLLIV
AHESPMSVFPWYSLSNNGYFKNISYFDIYGIELRLSNISTIIKSSPNIKLLSFSICFD
KLLYYIVKLNETQQKQQQQQQQQQLPFLQDKMIDDDNIIIEKCTCNCLLFYKSNINYK
KSLNCTSGGIDIDDEKQQSYNLNFVDHQEEEKEIQDDLNREINKFSNDWKVLCELLKN
SSIECLHLHQDCMDGLEIMQFNQLNRSSTTTKPNSIIYPDCFIQELGGSLISPINTLK
KISIMGLNTPQLFKYIVDHNKTIQNYIYTESILDLGYYNHQDYTSIFESIINENQQIT
ELKYSKAVFSLENEINMIDYEDYDADDDADELPYKDQVILHYKL"

CDS complement(13223. .14611)
/note="GeneID exon scores (in order of location ranges):
36.02 - GSCJ_ID dd_02813"
/codon_start=1
/product="similar to Plasmodium falciparum. Hypothetical
protein"
/protein_id="AAO53086.1"
/db_xref="GI:28850284"
/translation="MKVLSNLILKNIFLYLLNKNKFKQVLEYSLISKACFSIIQNLLT
FDNDCIMVKMTFNSIKSYIENSKLSINRPTQILNQLMKRKLLQFKEIEQLVIFYDNNN
NKNNNNENNDSNNDKELKRSEYNYLINNFKNLKRIYSESSKFAYFPSYQYYNDLVDYN
KICKIYIKKLYINWFYFGPISSSQEHQLYSNSPYEESIYFKFSFEFLSHYNPKKLIIS
TPNHQIQNNFYNLDNILNCKSIKSIYFKNIETPNSLIKKVFSETNIKSFKIILDNNNN
NQNNDNDNDNNGFILNRIENNNFKLKKLVIKNFNFYNNNNSKKEEGDINEQNKFLNEI
CLKNYNLKYIGLEFFVIKESNQLNQLFNLKFLNTISCSNDSFIAVIQHCNENKNIKTL
KYHNKVLVQLIPEDFKIFDSFFKSNKTLEDFRIFYYYPHLERSFLKVKKILSDSKNKT
IKNFEIKEVTKK"

CDS complement(join(14996. .16054,16142. .17131))
/note="GeneID exon scores (in order of location ranges):
89.73, 89.24 - GSCJ_ID dd_03199"
/codon_start=1
/product="similar to Dictyostelium discoideum (Slime
mold). Myb2 protein"
/protein_id="AAM45304.2"
/db_xref="GI:28850285"
/translation="MTAIFPNSNIPFYYWASTNAQPQQPQQSLQQQQQQQQQQQQQQQ
QQQQQQQHHHHQNIPPTPQPIFSPQLINKNFGCYGTSNQIPMIPQALFPQQTFNYSGL
PLCPSPPNYNNYHTITNSPPHQIHSPQLTIDQHSPPTCVSGELSPTPLSSSTGFSNNN
NNNNNNNNNNNNSNSNNNINNNNNNNIKRSYSAMMEENNYPNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNIINNNINSSGSISSYSNNSCIEDDYSSDKHCSDKESSNDCDDDAD
SYFEEYEDDQEAIRVSITPFVNLICKTPGITIPIIKIFSECSKEEITVKQMLLKIKNL
CLEFNLDNSIHNTLIHSLVKIDGDIKMVKYLIDNGISEQELEFTKDILLTSNSQEIIK
KKRERKRESISRGIRSPPNKWTKEESQNLIKLVTENGDKQWKKIATKLGGGKTGAQCA
QHWKRVLSPEIKKGSWDEAEEELLFQLVDKHGQSWKNVAIEIKTRTDIQCRYQYFKAI
MSRQTEWNQLEDDILTKKIKLMTQNNEKISFQQVSKHLARAKTTKIPRTALECKSRWS
QLNSTNVNNNNNNNNNSITTSSSNTNQQQQSTMVTPTSSPLSSPIPMSTPIINNSMPQ
QQVQQIQQQQISIQPYPQSFIQETQGFYRGNDQMQYHIQNQTIPQYHHQHNNNPLLML
"

CDS complement(join(18129. .18142,19312. .19449,22272. .22294,
22670. .22874,23418. .23622,23954. .24313,24384. .24563,
24644. .27214))
/note="GeneID exon scores (in order of location ranges):
0.17, 3.82, -0.39, 11.33, 8.29, 30.56, 7.20, 265.13 -
GSCJ_ID dd_00357"
/pseudo
/codon_start=1

CDS join(28837. .28849,28942. .29126)
/note="GeneID exon scores (in order of location ranges):
-2.77, 3.32 - GSCJ_ID dd_01206"

Query Match 21.0%; Score 47.2; DB 3; Length 85916;
Best Local Similarity 52.6%; Pred. No. 0.44;
Matches 103; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

```
Qy        30 TGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAGAATATTT 89
             ||   ||||  |  | | |      | |  ||| ||  ||||   ||       || ||||
Db     22208 TGTTTTAAACCACAATGTTATTATTTTTTTTATTTTAATTTTTATTTTTTTTAAAAATTT 22149

Qy        90 AGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAA 149
             | ||  |   |  ||||| | |   || || ||     |  ||   |  |||  || |||
Db     22148 ATTTATATAGAAATTTTTTATAAAAACTCTCTCAAATCACCAAGTTAGAAAATTAACAAA 22089
```

```
Qy       150 ACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAATCAATTA 209
             | ||   ||| ||||  ||    | |    ||| || | | |  | |  |  |  | |||
Db     22088 AAAAAAATTTAAAGATTTATAATCGAAGGTTTTTATTTACTTTTTTTTATTTAGTATTTA 22029

Qy       210 GGAATTTTTTTTTTTT 225
             ||||||||||||||||
Db     22028 ATAATTTTTTTTTTTT 22013
```

RESULT 7
AC094553
LOCUS AC094553 252248 bp DNA linear HTG 09-MAY-2003
DEFINITION Rattus norvegicus clone CH230-4P5, WORKING DRAFT SEQUENCE, 5 unordered pieces.
ACCESSION AC094553
VERSION AC094553.9 GI:30466688
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 252248)
AUTHORS Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M., Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E., Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J., Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L., Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K., Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L., Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H., Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D., Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K., Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V., Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 252248)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 252248)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 9, 2003 this sequence version replaced gi:24818672.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GAWO
                Center clone name: CH230-4P5
            -------------- Summary Statistics
                Assembly program: Atlas;
                Consensus quality: 225835 bases at least Q40
                Consensus quality: 228048 bases at least Q30
                Consensus quality: 229440 bases at least Q20
                Estimated insert size: 233224; sum-of-contigs estimation
                Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 5 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1   246259: contig of 246259 bp in length
            *  246260   246359: gap of unknown length
            *  246360   247679: contig of 1320 bp in length
            *  247680   247779: gap of unknown length
            *  247780   249446: contig of 1667 bp in length
            *  249447   249546: gap of unknown length
            *  249547   250637: contig of 1091 bp in length
            *  250638   250737: gap of unknown length
            *  250738   252248: contig of 1511 bp in length.
FEATURES             Location/Qualifiers
     source          1. .252248
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-4P5"
     misc_feature    1. .2282
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    7986. .8866
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH310954"
     misc_feature    240861. .242054
                     /note="wgs_contig"
     misc_feature    complement(240876. .241519)
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH310955"
     misc_feature    243149. .246259
                     /note="wgs_end_extension
                     clone_end:Sp6"
ORIGIN

  Query Match             20.5%;  Score 46.2;  DB 2;  Length 252248;
  Best Local Similarity   54.4%;  Pred. No. 0.62;
  Matches   93;  Conservative    0;  Mismatches   78;  Indels    0;  Gaps    0;

Qy         24 AAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAGA 83
              ||||||      || || |  |  |      | ||       ||   ||| ||   |||  |
Db     205591 AAAATTGCCTTCAATTTCTTTGTATAACAAAGATTGAGGCTTTTATTTTCCCATAGAGAA 205650

Qy         84 ATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAAC 143
                 || | ||||||| ||  |||||||     |    || |   ||||||  ||| ||||
Db     205651 TGCTTGAATTTCACTAAGAATTTTCAAGTACATTTAGTGCTGAAAAAAAAAAAAGAAAAG 205710

Qy        144 AAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCT 194
              |||| || ||      ||   ||||||| | |||| |||     |  ||| |
Db     205711 AAAAGAAGAAAGAAATTTCCAGTTGATTGTTACTAATTAAGAATTTCTCTT 205761


RESULT 8
PFMAL13_04
WPCOMMENT
  Sequence split into 28 fragments  LOCUS PFMAL13 Accession AL844509
     Fragment Name      Begin     End
     PFMAL13_00             1    110000
     PFMAL13_01        100001    210000
     PFMAL13_02        200001    310000
     PFMAL13_03        300001    410000
     PFMAL13_04        400001    510000
     PFMAL13_05        500001    610000
     PFMAL13_06        600001    710000
     PFMAL13_07        700001    810000
     PFMAL13_08        800001    910000
     PFMAL13_09        900001   1010000
     PFMAL13_10       1000001   1110000
     PFMAL13_11       1100001   1210000
     PFMAL13_12       1200001   1310000
     PFMAL13_13       1300001   1410000
     PFMAL13_14       1400001   1510000
     PFMAL13_15       1500001   1610000
     PFMAL13_16       1600001   1710000
     PFMAL13_17       1700001   1810000
     PFMAL13_18       1800001   1910000
     PFMAL13_19       1900001   2010000
     PFMAL13_20       2000001   2110000
     PFMAL13_21       2100001   2210000
     PFMAL13_22       2200001   2310000
     PFMAL13_23       2300001   2410000
     PFMAL13_24       2400001   2510000
     PFMAL13_25       2500001   2610000
     PFMAL13_26       2600001   2710000
     PFMAL13_27       2700001   2732359
Continuation (5 of 28) of PFMAL13 from base 400001 (AL844509 Plasmodium falciparum 3D7 c

  Query Match             20.4%;  Score 45.8;  DB 2;  Length 110000;
  Best Local Similarity   51.2%;  Pred. No. 0.87;
  Matches  107;  Conservative    0;  Mismatches  102;  Indels    0;  Gaps    0;

Qy         17 CAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTT 76
                ||| |||||||||  ||    |||     | ||    |   ||| |   |  ||  |
```

```
Db      28080 CAGCTATAAAATTTACAACGTCCTATTTATATACACATAAATATATGTATATAATATATA 28139

Qy         77 CAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAA 136
              | |   |||||||  |  |    |  | |||  || |||           ||||||  ||
Db      28140 TATATATATATTTATATATATAAAAATATTTTGAAACTAGCGAAAAAAAAAAAAAAAAAA 28199

Qy        137 ACAAAACAAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCT 196
              | ||||  |||||||  | |||||||| |  |  |  ||     ||| || |   |
Db      28200 AAAAAAAAAAAAAAAATAGCTTTTTAAAATATATTTCCTTGCTGTTTCATGTATTTTTAAA 28259

Qy        197 GCTGAATCAATTAGGAATTTTTTTTTTT 225
                | |   |  |     ||||||||||||
Db      28260 AATAAGAAATATTTTTTTTTTTTTTTTT 28288


RESULT 9
U67577/c
LOCUS       U67577                 9971 bp    DNA     linear   BCT 28-JAN-1998
DEFINITION  Methanococcus jannaschii section 119 of 150 of the complete genome.
ACCESSION   U67577 L77117
VERSION     U67577.1  GI:2826400
KEYWORDS    .
SOURCE      Methanocaldococcus jannaschii
  ORGANISM  Methanocaldococcus jannaschii
            Archaea; Euryarchaeota; Methanococci; Methanococcales;
            Methanocaldococcaceae; Methanocaldococcus.
REFERENCE   1  (bases 1 to 9971)
  AUTHORS   Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
            Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
            Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J., Adams,M.D.,
            Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
            Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,
            Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M.,
            Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
            Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,
            Smith,H.O., Woese,C.R. and Venter,J.C.
  TITLE     Complete genome sequence of the methanogenic archaeon,
            Methanococcus jannaschii
  JOURNAL   Science 273 (5278), 1058-1073 (1996)
  MEDLINE   96337999
   PUBMED   8688087
REFERENCE   2  (bases 1 to 9971)
  AUTHORS   Bult,C.J., White,O., Olsen,G.J., Zhou,L., Fleischmann,R.D.,
            Sutton,G.G., Blake,J.A., FitzGerald,L.M., Clayton,R.A.,
            Gocayne,J.D., Kerlavage,A.R., Dougherty,B.A., Tomb,J.-F.,
            Adams,M.D., Reich,C.I., Overbeek,R., Kirkness,E.F., Weinstock,K.G.,
            Merrick,J.M., Glodek,A., Scott,J.D., Geoghagen,N.S., Weidman,J.F.,
            Fuhrmann,J.L., Nguyen,D.T., Utterback,T., Kelley,J.M.,
            Peterson,J.D., Sadow,P.W., Hanna,M.C., Cotton,M.D., Hurst,M.A.,
            Roberts,K.M., Kaine,B.B., Borodovsky,M., Klenk,H.P., Fraser,C.M.,
            Smith,H.O., Woese,C.R. and Venter,J.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (27-AUG-1996) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
COMMENT     On Jan 30, 1998 this sequence version replaced gi:1592013.
FEATURES             Location/Qualifiers
     source          1. .9971
                     /organism="Methanocaldococcus jannaschii"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:2190"
     gene            87. .554
                     /gene="MJ1369"
     CDS             87. .554
                     /gene="MJ1369"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJ1369"
                     /protein_id="AAB99386.1"
                     /db_xref="GI:1592014"
                     /translation="MRDIVMTYWLFSSNNIRNIEICYNHMIWGFWDRNAGEKQKKNWR
                     SFIRKYNQIKPFDVAVFQIAKTGEIHAIGVIKETYYDDQTPIWDNEINLNKVTFPWRV
                     SFSVIIFSKEAVIKRFIKIQDYIDGYGLGELEHHDFNEILKAFQKKFGMISIK"
     gene            584. .1570
                     /gene="MJ1370"
     CDS             584. .1570
                     /gene="MJ1370"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJ1370"
                     /protein_id="AAB99387.1"
                     /db_xref="GI:1592015"
                     /translation="MITVGIDHGTSGITTCIKDNDKKIIFKLKRTELKEKSYLEELEK
                     HISLEDIDLIALTYSMGDGINKILPIEKVKNRGVLSIEGAGEKVGGGTKVYDEIKESG
                     LPAVVIPGLHRGIECLDERFRALYSHIASPEKVSIAYYAYKLFGFNDFVLSDISSNTV
                     TLLIKDGKIFGGFDACIGAIGMLHGPIDLEMIRDIDAGKITANEAFSKAGAVKIAKLY
                     KGVENTKEEIINNYFNDENCRLAVDSLILSVSMEINSLLPLLDKNKRRVVLAGSIGTL
                     RNPIDIPKRIKEFVEAKIFVLYGESGAIGGALIAEDILKGKRDILGIEVEFK"
     gene            1567. .2301
                     /gene="MJ1371"
     CDS             1567. .2301
                     /gene="MJ1371"
                     /note="similar to GP:2437836 percent identity: 39.09;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="methlytransferase"
                     /protein_id="AAB99379.1"
                     /db_xref="GI:1592016"
                     /translation="MIIMYAIGIGDNKEEVLKAYEKLKEEGIEVELIDNPKLLVDKLL
                     DGEIDGAVRGSLSSSKVILYLRERIGKFYRASILKNPFTNGIFLLSPVGIDDISEDKN
                     ERIKDKIRIIEFASNFLKNYNIKAKVAVLSGGRLGDLGRNKVVDETIYEAEEIVEHFK
                     GNVDIIHNGILIEEYLKDGYNIIIAVDGITGNLIFRCLGLICKIPGYGAVILSDKNVN
                     FIDTSRNANWERYYNAIKFLIGGDFG"
     gene            2285. .3127
                     /gene="MJ1372"
     CDS             2285. .3127
                     /gene="MJ1372"
                     /note="similar to GB:L42023 SP:P44938 PID:1006041
                     PID:1221025 PID:1205167 percent identity: 47.86;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAB99380.1"
                     /db_xref="GI:1592017"
                     /translation="MGILGKIKNKLKSIGKRVIIDFYRFLDNSGVLKIYEKILEEAID
                     KDNLPKHVAIIMDGNRRAAEIYGKDRYYGHYLGAEKVREVLRWARDLGINVVTLYAFS
                     TENFRRPKEEVDKLMELFEKKFYEIADDEEIHRYEVRVRAIGRINLLPKNVQKAIKYA
                     EERTKNYNKFFVNIAIAYGGQQEIIDAVKKIAEKVKRGEIEPEDIDKELIDKHLYTAN
                     LPFPNPDLIIRTSGEERISNFLIWQSSYSELYFCDIYWPLFRRVDFLRAVRDYQRRQR
                     RFGK"
     gene            3124. .3612
                     /gene="MJ1373"
     CDS             3124. .3612
                     /gene="MJ1373"
                     /note="hypothetical protein; identified by GeneMark;
                     putative"
                     /codon_start=1
                     /transl_table=11
                     /product="M. jannaschii predicted coding region MJ1373"
                     /protein_id="AAB99388.1"
                     /db_xref="GI:1592018"
                     /translation="MKKKIIILFLFTAILCSITLCGCISFEKTQIGNYNIKNNSCINS
                     NVSKNQTIQNVSDKIVNNNTNILNTLNYEIIAYGAFGEKNRGYYYYYKDNKTIIVINL
                     EEMPTAGYKIKIINITETANKITVYYKVIPPKEFAAMVVTYPYIKLSVNGTYNVECKE
                     VR"
     gene            3680. .4480
                     /gene="MJ1374"
     CDS             3680. .4480
                     /gene="MJ1374"
                     /note="similar to GB:AE000666 percent identity: 42.73;
                     identified by sequence similarity; putative"
```

```
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAB99381.1"
                     /db_xref="GI:1592019"
                     /translation="MRVEIIFLGCGGGRWATITQKKATGGFRIHTNELRMHVDPGPGA
                     IVRLNELKISPWRTNALFISHCHPDHYTDGEIIVEAITQGMTKKRGVFLGSLSVVEGF
                     GEYEYVVSKYHQSKLEEVRVLYPGDSAELYDTTIKATHTKHGDPFGIGFRLSTIYGDI
                     GYTSDTEFIPQLIEDFDGVRILIANIVRKKNERIKGHLCSNDAIDLINSMNKKPELLI
                     MNHMGVKMTNPQIEAEYISQNTGIEVIPARLGLKVELLNGKYKYQLIK"
gene                 complement(4473. .5720)
                     /gene="MJ1375"
CDS                  complement(4473. .5720)
                     /gene="MJ1375"
                     /note="similar to GB:D26185 SP:P37555 PID:467446
                     GB:AL009126 percent identity: 23.61; identified by
                     sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="SpoVB isolog"
                     /protein_id="AAB99382.1"
                     /db_xref="GI:2826401"
                     /translation="MSLCKDSIYILMSNLYSKGMAYLFYFITAFLLGTEAFGILKGLM
                     PIADTLTIFFSSGIPPAIAKFLAEEKEVDINKYIPILYLMILLSVVGFILTPYIKYIL
                     GGHYLNLPNILYFAVGLCVVASTVIAFSRGILQGLLKMKYLSLTWIVEYTAKVILVFI
                     LTLYLGIFGSLLSISLAYLVGGIFGLYLIYKALKGKFDFKKLIDIKNTTKNIFSNFNL
                     DILRYSIPIALTSSSYRLFGDIDNIVIMSIMGGFWSGIYGYSSLISRGIFMFASAVSI
                     PLLPRISKTKDLSLLKEGIIQNTIFSSIFVIGCLFFPEIPLIAFFKTANPEGILCLRI
                     LAISSLFMSYYTLISSALQGLGYAKISFYIILFGLVLNIILNLILVNAYGIVGGSLAT
                     LITSISVFLIGVFAILRIKKHII"
gene                 5807. .6544
                     /gene="MJ1376"
CDS                  5807. .6544
                     /gene="MJ1376"
                     /note="similar to GB:L42023 SP:P45162 PID:1007333
                     PID:1221464 PID:1205570 percent identity: 40.10;
                     identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="cell division protein FtsJ"
                     /protein_id="AAB99383.1"
                     /db_xref="GI:1592021"
                     /translation="MGRKDKRWVLQRKRDFYYKLAKKLKYRSRASFKLMQLNEKFNVI
                     KPGKIVLDLGCAPGGWMQVAREIVGDKGFVIGIDLQPVKPFEYDNVVAIKGDFTLEEN
                     LNKIRELIPNDEKKVDVVISDASPNISGYWDIDHARSIDLVTTALQIATEMLKERGNF
                     VAKVFYGDMIDDYVNLVKKYFEKVYITKPQASRKESAEVYVIAKRYTGKKWEEEDKIK
                     RVKKVENEDNELLAKKIKEIRKLKSKK"
gene                 complement(6556. .8217)
                     /gene="MJ1377"
CDS                  complement(6556. .8217)
                     /gene="MJ1377"
                     /note="similar to PID:1051202 SP:Q50543 percent identity:
                     51.36; identified by sequence similarity; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="glutamyl-tRNA synthetase (gltX)"
                     /protein_id="AAB99384.1"
                     /db_xref="GI:1592022"
                     /translation="MEEKILPIALRNAIKYNGKANPKAVLGIFLSENPEYRSKAKEVM
                     PIVEKVVEEVNKLSLDEIKKKLEELGEDVKKKEKKEKGLELPNVKDKVVMRFAPNPSG
                     PLHIGHARAAVLNDYFVKKYGGKLILRLEDTDPKRVLPEAYDMIKEDLDWLGVKVDEV
                     VIQSDRIELYYEYGRKLIEMGHAYVCDCNPEEFRELRNKGVPCKCRDRAIEDNLELWE
                     KMLNGELENVAVRLKTDIKHKNPSIRDFPIFRVEKTPHPRTGDKYCVYPLMNFSVPVD
                     DHLLGMTHVLRGKDHIVNTEKQAYIYKYFGWEMPEFIHYGILKIEDIVLSTSSMYKGI
                     KEGLYSGWDDVRLGTLRALRRRGIKPEAIYEIMKRIGIKQADVKFSWENLYAINKELI
                     DKDARRFFFVWNPKKLIIEGAEKKVLKLRMHPDRPEFGERELIFDGEVYVVGDELEEN
                     KMYRLMELFNIVVEKVDDIALAKYHSDDFKIARKNKAKIIHWIPVKDSVKVKVLMPDG
                     EIKEGFAEKDFAKVEVDDIIQFERFGFVRIDKKDNDGFVCCYAHR"
gene                 8402. .9850
                     /gene="MJ1378"
CDS                  8402. .9850

Query Match          20.1%;  Score 45.2;  DB 1;  Length 9971;
  Best Local Similarity   51.0%;  Pred. No. 1.8;
  Matches  107;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          7 TAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTG 66
              ||| |   | ||| | |   |||| |   | | |   ||| |||| | || || ||||
Db       2481 TAAATCTCTGCAGCCCTTCTATTTCCATCCATTATTATAGCTACATGTTTTGGTAAATTG 2422

Qy         67 GTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCC 126
              |||| |   ||      | | | | |          |||| ||| || || ||   |
Db       2421 TCTTTATCAATAGCCTCTTCTAAAATCTTCTCGTAAATTTTTAAAACTCCGGAGTTGTCT 2362

Qy        127 AAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATC 186
              ||||| | | | ||| ||| |    | |||| |   |   | |    ||| |||
Db       2361 AAAAATCTATAAAAATCAATTATTACTCTTTTTCCAATACTCTTTAATTTGTTTTTTATC 2302

Qy        187 TGCCTCCTCTGCTGAATCAATTAGGAATTT 216
              | | |     |   | | ||||||||||||
Db       2301 TTACCCAAAATCCCCACCTATTAGGAATTT 2272


RESULT 10
AR271569_01
WPCOMMENT
  Sequence split into 17 fragments  LOCUS AR271569 Accession AR271569
     Fragment Name      Begin     End
     AR271569_00            1    110000
     AR271569_01       100001    210000
     AR271569_02       200001    310000
     AR271569_03       300001    410000
     AR271569_04       400001    510000
     AR271569_05       500001    610000
     AR271569_06       600001    710000
     AR271569_07       700001    810000
     AR271569_08       800001    910000
     AR271569_09       900001   1010000
     AR271569_10      1000001   1110000
     AR271569_11      1100001   1210000
     AR271569_12      1200001   1310000
     AR271569_13      1300001   1410000
     AR271569_14      1400001   1510000
     AR271569_15      1500001   1610000
     AR271569_16      1600001   1664976
Continuation (2 of 17) of AR271569 from base 100001 (AR271569 Sequence 1 from patent US

  Query Match          20.1%;  Score 45.2;  DB 6;  Length 110000;
  Best Local Similarity   51.0%;  Pred. No. 1.2;
  Matches  107;  Conservative    0;  Mismatches  103;  Indels    0;  Gaps    0;

Qy          7 TAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTG 66
              ||| |   | ||| | |   |||| |   | | |   ||| |||| | || || ||||
Db      51403 TAAATCTCTGCAGCCCTTCTATTTCCATCCATTATTATAGCTACATGTTTTGGTAAATTG 51462

Qy         67 GTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCC 126
              |||| |   ||      | | | | |          |||| ||| || || ||   |
Db      51463 TCTTTATCAATAGCCTCTTCTAAAATCTTCTCGTAAATTTTTAAAACTCCGGAGTTGTCT 51522

Qy        127 AAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATC 186
              ||||| | | | ||| ||| |    | |||| |   |   | |    ||| |||
Db      51523 AAAAATCTATAAAAATCAATTATTACTCTTTTTCCAATACTCTTTAATTTGTTTTTTATC 51582

Qy        187 TGCCTCCTCTGCTGAATCAATTAGGAATTT 216
              | | |     |   | | ||||||||||||
Db      51583 TTACCCAAAATCCCCACCTATTAGGAATTT 51612


RESULT 11
AC096278
LOCUS       AC096278              231461 bp    DNA     linear   HTG 10-MAY-2003
DEFINITION  Rattus norvegicus clone CH230-11L24, WORKING DRAFT SEQUENCE.
ACCESSION   AC096278
VERSION     AC096278.5  GI:30522073
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
```

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 231461)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Loulseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 231461)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-SEP-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 231461)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 10, 2003 this sequence version replaced gi:23267195.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GEVR
                Center clone name: CH230-11L24
            -------------- Summary Statistics
                Assembly program: Atlas 3.0;
                Consensus quality: 225324 bases at least Q40
                Consensus quality: 227015 bases at least Q30
                Consensus quality: 227940 bases at least Q20
                Estimated insert size: 232676; sum-of-contigs estimation
                Quality coverage: 8x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1   231461: contig of 231461 bp in length.
FEATURES              Location/Qualifiers
     source           1. .231461
                      /organism="Rattus norvegicus"
                      /mol_type="genomic DNA"
                      /db_xref="taxon:10116"
                      /clone="CH230-11L24"
     misc_feature     1. .1012
                      /note="wgs_end_extension
                      clone_end:T7"
     misc_feature     2038. .2914
                      /note="clone_boundary
                      clone_end:T7
                      site:EcoRI
                      end_sequence:BH343611"
     misc_feature     228529. .229317
                      /note="clone_boundary
                      clone_end:Sp6
                      site:EcoRI
                      end_sequence:BH343613"
     misc_feature     230067. .231461
                      /note="wgs_end_extension
                      clone_end:Sp6"
ORIGIN

  Query Match             20.1%;  Score 45.2;  DB 2;  Length 231461;
  Best Local Similarity   53.2%;  Pred. No. 1;
  Matches  118;  Conservative    0;  Mismatches  103;  Indels    1;  Gaps    1;

Qy          3 ATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTAT 62
              |||  || || |   ||||||  || | | |||  ||||||  |   ||  |  |||   ||
Db     150885 ATGAGAATTTTTAATAGGCATCAAGTATTAAACCAATTATTATTTTCAGTTTATAGAAT 150944

Qy         63 ATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTC 122
               |    | ||  || | ||    |     |||   | |   || || |||  | | |
Db     150945 GTGATGTATAATTTAATAATGTAACAATATTTTGTTTACCATTGTCTAAGGAAAGATTCT 151004

Qy        123 CCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTT 182
```

```
                ||  ||||| |||       |||||  || |  |||||   | || |||      ||| |
Db      151005 TCCTCAAAAATGAATTTTTCAAAAAATACTAAATTTTTTTCACTT-ATGTAAGTACATCT 151063

Qy         183 GATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTT 224
               | || || ||    ||| | | ||   |||||| | || |
Db      151064 CTTTTGGCTGCTTGATTGATTTATTTCCAAATTTTCTATTAT 151105
```

RESULT 12
AC146265/c
LOCUS AC146265 170419 bp DNA linear PRI 29-OCT-2003
DEFINITION Pan troglodytes BAC clone RP43-28O21 from 7, complete sequence.
ACCESSION AC146265
VERSION AC146265.2 GI:37574299
KEYWORDS HTG.
SOURCE Pan troglodytes (chimpanzee)
ORGANISM Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE 1 (bases 1 to 170419)
AUTHORS Van Brunt,A. and Haglund,K.
TITLE The sequence of Pan troglodytes BAC clone RP43-28O21
JOURNAL Unpublished (2001)
REFERENCE 2 (bases 1 to 170419)
AUTHORS Sulston,J.E. and Wilson,R.
TITLE Sequencing of Pan troglodytes
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 170419)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (01-AUG-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 4 (bases 1 to 170419)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (10-SEP-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 170419)
AUTHORS Wilson,R.K.
TITLE Direct Submission
JOURNAL Submitted (08-OCT-2003) Genetics, Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 6 (bases 1 to 170419)
AUTHORS Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (29-OCT-2003) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Oct 8, 2003 this sequence version replaced gi:33387218.
---------- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@watson.wustl.edu
---------- Summary Statistics
Center project name: C_PT028O21
----------.

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-43 BAC Library has been constructed by Chung-Li Shu. DNA was isolated from white blood cells obtained from a male chimpanzee (Pan troglodytes,'Clint', Yerkes #C0471; birthdate: 6-6-80). The clone and detailed information can be obtained from ResGen (http://www.resgen.com) or Pieter de Jong and co-workers at http://www.bacpac.chori.org.

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone.

FEATURES Location/Qualifiers
source 1. .170419
/organism="Pan troglodytes"
/mol_type="genomic DNA"
/db_xref="taxon:9598"
/chromosome="7"
/map="7"
/clone="RP43-28O21"
/clone_lib="RPCI-43"
unsure 46167. .46239
/note="Sequence derived from one plasmid subclone."
unsure 46401. .46732
/note="Sequence derived from one plasmid subclone."
misc_feature 46821. .47724
/note="Sequence derived from PCR product of project DNA."
unsure 73602. .73832
/note="Sequence derived from one plasmid subclone."
misc_feature 75265. .75545
/note="Sequence derived from PCR product of project DNA."
unsure 143618. .143701
/note="Sequence derived from one plasmid subclone."
repeat_region 169962. .169963
/transposon="Bacterial transposon insertion in clone excised here."

ORIGIN

Query Match 20.0%; Score 45; DB 9; Length 170419;
Best Local Similarity 54.5%; Pred. No. 1.2;
Matches 90; Conservative 0; Mismatches 75; Indels 0; Gaps 0;

```
Qy          55 TATGCTATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCT 114
               ||   | | |||| |||  ||||       | ||||  ||   ||    |||||   | |
Db       40410 TACAGTTTTTTGGGTTTGGCTTCCTTCTCACATTTCTTTAGCCTTGAATTTTTACTAAAT 40351

Qy         115 ACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCT 174
               |   | | ||  |  ||  ||| |||| |||||||  |    ||| ||| || | ||
Db       40350 AAATTTCCTCCTGATTAAAAAAAAAAAAAAAAAAAAAAAGAGGAGTTTTAGATGTGGTCCCT 40291

Qy         175 ACTCATTTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTT 219
               ||  ||||      |||| |    | ||||| ||  | ||||
Db       40290 TTTCTATTGAATGTGGTCCTTTTTCTATTCAATTTGGTCTCTTTT 40246
```

RESULT 13
AL929001/c
LOCUS AL929001 147727 bp DNA linear ROD 16-JAN-2003
DEFINITION Mouse DNA sequence from clone RP23-324H1 on chromosome 2, complete sequence.
ACCESSION AL929001
VERSION AL929001.7 GI:27801672
KEYWORDS HTG.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 147727)
AUTHORS Almeida,J.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries:

```
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Jan 17, 2003 this sequence version replaced gi:27764086.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence.  Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            --------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-324H1 is
            from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES                Location/Qualifiers
     source             1. .147727
                        /organism="Mus musculus"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:10090"
                        /chromosome="2"
                        /clone="RP23-324H1"
                        /clone_lib="RPCI-23"
ORIGIN

  Query Match             19.7%;  Score 44.4;  DB 10;  Length 147727;
  Best Local Similarity   62.7%;  Pred. No. 1.7;
  Matches   69;  Conservative    0;  Mismatches   41;  Indels    0;  Gaps    0;

Qy          74 CTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAAC 133
               || || |   |  |  ||   ||||  || | | || ||  || ||||| |  ||||||
Db       87490 CTACAAAGTGAGTTCCAGGACCACTAGGGCTATACAGAGAAACACTGTCTCGGAAAAAAC 87431

Qy         134 GAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTG 183
               ||||||||||||| ||||| | ||     || ||    |||   |||||
Db       87430 AAAACAAAACAAAACAACAAACATTCATTTAGGTGGATTCTAAGAATTTG 87381


RESULT 14
AC104893/c
LOCUS       AC104893               148750 bp    DNA     linear   HTG 16-JUL-2003
DEFINITION  Mus musculus clone RP23-288O15, WORKING DRAFT SEQUENCE, 5 ordered
            pieces.
ACCESSION   AC104893
VERSION     AC104893.4  GI:32813580
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 148750)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP23-288O15
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 148750)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-DEC-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 148750)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 16, 2003 this sequence version replaced gi:20043160.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L19047
                Center clone name: 288_O_15
            -------------- Summary Statistics
                Sequencing vector: Plasmid; n/a; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 147390 bases at least Q40
                Consensus quality: 147957 bases at least Q30
                Consensus quality: 148226 bases at least Q20
                Insert size: 152000; agarose-fp
                Insert size: 148350; sum-of-contigs
                Quality coverage: 9.7 in Q20 bases; agarose-fp
```

```
            Quality coverage: 9.9 in Q20 bases; sum-of-contigs
           --------------.
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 5 contigs. Gaps between the contigs
           * are represented as runs of N. The order of the pieces
           * is believed to be correct as given, however the sizes
           * of the gaps between them are based on estimates that have
           * provided by the submittor.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *        1    67157: contig of 67157 bp in length
           *    67158    67257: gap of 100 bp
           *    67258    70312: contig of 3055 bp in length
           *    70313    70412: gap of 100 bp
           *    70413    76160: contig of 5748 bp in length
           *    76161    76260: gap of 100 bp
           *    76261    91722: contig of 15462 bp in length
           *    91723    91822: gap of 100 bp
           *    91823   148750: contig of 56928 bp in length.
FEATURES             Location/Qualifiers
     source          1. .148750
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-288O15"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1. .67157
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    67258. .70312
                     /note="assembly_fragment"
     misc_feature    70413. .76160
                     /note="assembly_fragment"
     misc_feature    76261. .91722
                     /note="assembly_fragment"
     misc_feature    91823. .148750
                     /note="assembly_fragment"
ORIGIN

  Query Match             19.6%;  Score 44;  DB 2;  Length 148750;
  Best Local Similarity   51.5%;  Pred. No. 2.1;
  Matches  101;  Conservative    0;  Mismatches   95;  Indels    0;  Gaps    0;

Qy         23 TAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAG 82
              ||||| |||||||||  |  |   | ||   | | | | ||        | ||| | ||
Db      97863 TAAAAAATGAAATAAAAAAAAATCATACACATTTTTCCAAATAAAGCACATCTTTATAAA 97804

Qy         83 AATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAA 142
              || ||||||   || ||||| || |  |  |  |    | | |||||   | |  ||
Db      97803 AAAATTTAGGAACAGTCAGGGTTATACAGAGAAACCCTGTCTCAAAAAACAAACAACAAC 97744

Qy        143 CAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGAA 202
              | || || || |  |||||    ||  |  ||     |   |||   |||| |  ||
Db      97743 AACAACAAAAATCAATTTAATTTATGCAGAATAGCTCCTATTTCTTTATCCTATCATGCT 97684

Qy        203 TCAATTAGGAATTTTT 218
              |   |   ||||||
Db      97683 AAATAAAATTATTTTT 97668


RESULT 15
AC117237/c
LOCUS       AC117237               200110 bp    DNA     linear   ROD 05-NOV-2003
DEFINITION  Mus musculus BAC clone RP23-431G18 from 13, complete sequence.
ACCESSION   AC117237
VERSION     AC117237.2  GI:21206431
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 200110)
  AUTHORS   Radionenko,M., Bielicki,L. and Doebber,A.
  TITLE     The sequence of Mus musculus BAC clone RP23-431G18
  JOURNAL   Unpublished (2001)
REFERENCE   2  (bases 1 to 200110)
  AUTHORS   Wilson,R.
  TITLE     Sequencing of Mus musculus
  JOURNAL   Unpublished (2001)
REFERENCE   3  (bases 1 to 200110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4  (bases 1 to 200110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-MAY-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5  (bases 1 to 200110)
  AUTHORS   McPherson,J.D. and Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6  (bases 1 to 200110)
  AUTHORS   Wilson,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On May 25, 2002 this sequence version replaced gi:20069804.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu
                Contact: submissions@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: M_BA0431G18
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            Mapping information for this clone was provided by Dr. Wes Warren,
            Department of Genetics, Washington University, St. Louis MO. For
            additional information about the map position of this sequence, see
            http://genome.wustl.edu

            SOURCE INFORMATION:
            The RPCI-23 BAC Library has been constructed by Kazutoyo Osegawa
            and Minako Tateno in the laboratory of Pieter de Jong
            (http://www.chori.org) from female C57BL/6J mouse kidney and/or
            brain genomic DNA.  The clone and detailed information can be
            obtained from Research Genetics, Inc. (http://www.resgen.com) or
            Pieter de Jong and coworkers at http://www.chori.org

            NEIGHBORING SEQUENCE INFORMATION:
            This sequence is the entire insert of the clone.
FEATURES             Location/Qualifiers
     source          1. .200110
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="13"
```

```
                     /map="13"
                     /clone="RP23-431G18"
                     /clone_lib="RPCI-23"
repeat_region        1081. .1426
                     /rpt_family="MaLR"
repeat_region        1909. .3410
                     /rpt_family="L1"
repeat_region        4998. .5172
                     /rpt_family="RMER15"
repeat_region        5225. .5382
                     /rpt_family="RMER15"
repeat_region        5394. .5449
                     /rpt_family="RMER15"
repeat_region        6029. .6271
                     /rpt_family="MaLR"
repeat_region        6475. .6679
                     /rpt_family="MaLR"
repeat_region        6898. .7087
                     /rpt_family="B4"
repeat_region        7621. .7703
                     /rpt_family="L2"
repeat_region        7726. .7907
                     /rpt_family="B2"
repeat_region        8112. .8387
                     /rpt_family="B4"
repeat_region        8708. .8882
                     /rpt_family="L1"
repeat_region        8898. .9398
                     /rpt_family="L1"
repeat_region        10449. .10511
                     /rpt_family="B2"
repeat_region        10512. .10691
                     /rpt_family="L1"
repeat_region        10686. .11044
                     /rpt_family="L1"
repeat_region        11068. .11125
                     /rpt_family="B2"
repeat_region        11125. .11327
                     /rpt_family="B2"
repeat_region        12120. .12239
                     /rpt_family="Alu"
repeat_region        12248. .12396
                     /rpt_family="B4"
repeat_region        12275. .12398
                     /rpt_family="Alu"
repeat_region        13267. .13467
                     /rpt_family="B2"
repeat_region        13926. .14052
                     /rpt_family="MIR"
repeat_region        14980. .15162
                     /rpt_family="B2"
repeat_region        15636. .15735
                     /rpt_family="BC1_MM"
repeat_region        15811. .15960
                     /rpt_family="B4"
repeat_region        17133. .17177
                     /rpt_family="RMER10A"
repeat_region        18551. .18947
                     /rpt_family="MaLR"
repeat_region        21645. .21701
                     /rpt_family="L1"
repeat_region        21702. .21777
                     /rpt_family="L1"
repeat_region        21778. .22099
                     /rpt_family="MaLR"
repeat_region        22100. .22386
                     /rpt_family="L1"
repeat_region        23469. .23542
                     /rpt_family="L2"
repeat_region        24176. .24549
                     /rpt_family="MaLR"
repeat_region        26163. .26303
                     /rpt_family="Alu"
repeat_region        26542. .27074
                     /rpt_family="L1"
repeat_region        29252. .29294
                     /rpt_family="MER105"
repeat_region        29355. .30206
                     /rpt_family="L1"
repeat_region        30847. .31034
                     /rpt_family="B2"
repeat_region        31524. .31657
                     /rpt_family="L1"
repeat_region        31831. .31966
                     /rpt_family="B4"
repeat_region        32682. .33406
                     /rpt_family="L1"
repeat_region        33515. .33604
                     /rpt_family="L1"
repeat_region        33680. .33779
                     /rpt_family="B4"
repeat_region        33793. .34073
                     /rpt_family="L1"
repeat_region        34320. .34790
                     /rpt_family="L1"
repeat_region        34794. .34941
                     /rpt_family="L1"
repeat_region        35576. .36109
                     /rpt_family="RMER15"
repeat_region        36349. .36432
                     /rpt_family="B2"
repeat_region        36452. .36669
                     /rpt_family="MER1_type"
repeat_region        38048. .38444
                     /rpt_family="RMER6B"
repeat_region        38753. .38967
                     /rpt_family="L2"
repeat_region        39216. .39367
                     /rpt_family="RMER19B"
repeat_region        39664. .40186
                     /rpt_family="ERV1"
repeat_region        41800. .41948
                     /rpt_family="MER1_type"
repeat_region        42216. .42473
                     /rpt_family="L1"
repeat_region        42469. .42786
                     /rpt_family="L1"
repeat_region        44614. .44761

 Query Match           19.5%;  Score 43.8;  DB 10;  Length 200110;
 Best Local Similarity  55.6%;  Pred. No. 2.2;
 Matches   84;  Conservative    0;  Mismatches   67;  Indels    0;  Gaps    0;

Qy         75 TTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACG 134
              | ||||    |  |  ||      |    || |   || ||  || |||||  |  ||||||
Db     105570 TACAGAGTGAGTTCCAGGACAGCCAGGGCTACACAGAGAAACCCTGTCTCAAAAAAAAAA 105511

Qy        135 AAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCT 194
              ||| |||| ||||||||||||    |    |||||||  ||   ||  ||||   |
Db     105510 AAAAAAAAAAAAAAAAACAACCAGAAAATATTTTGATGGGAACAGTTTATATCTTAATTTA 105451

Qy        195 CTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |  |   |  ||||||  || | | | |||
Db     105450 TTTATTTTTACATTAGGTGTTATGTGTGTTT 105420


Search completed: July 27, 2004, 19:21:01
Job time : 3661 secs
```

GenCore version 5.1.6

OM nucleic - nucleic search, using sw model

Run on: July 27, 2004, 17:01:40 ; Search time 407 Seconds
(without alignments)
2348.512 Million cell updates/sec

Title: US-09-765-231A-58
Perfect score: 225
Sequence: 1 tgatggtaagttgtttcagg.....attaggaatttttttttttt 225

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_29Jan04:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002s:*
7: geneseqn2003as:*
8: geneseqn2003bs:*
9: geneseqn2003cs:*
10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	225	100.0	225	4	AAH23128	Aah23128 Osteoarth
	2	225	100.0	320	7	ACA04823	Aca04823 cDNA enco
	3	48.4	21.5	110000	2	AAV21209_01	Continuation (2 of
c	4	41	18.2	5518	6	ABK28306	Abk28306 DNA trans
c	5	40	17.8	6809	6	ABK31314	Abk31314 Signal tr
c	6	40	17.8	6809	6	ABL70557	Abl70557 Chemicall
c	7	40	17.8	6809	6	AAS61214	Aas61214 Human gen
c	8	40	17.8	6809	6	ABN80174	Abn80174 Human che
c	9	38.2	17.0	110000	6	ABA92787_1	Continuation (2 of
c	10	38	16.9	12393	6	ABL33263	Abl33263 Human imm
c	11	37.8	16.8	96588	8	ADA03026	Ada03026 Human MBN
c	12	37.8	16.8	96588	9	ADB72764	Adb72764 Human MBN
c	13	37.8	16.8	96588	9	ADC85506	Adc85506 Human Mbn
c	14	37.4	16.6	2270	2	AAX05715	Aax05715 Human pro
c	15	37.2	16.5	778	4	AAL21862	Aal21862 Human bre
	16	36.6	16.3	2705	2	AAV05164	Aav05164 Human gro
c	17	36.4	16.2	345	6	ABQ85902	Abq85902 Arabidops
	18	36.4	16.2	1479	6	ABN95163	Abn95163 Gene #166
c	19	36.4	16.2	7924	6	ABK40070	Abk40070 Human che
c	20	36.4	16.2	7924	6	ABL34139	Abl34139 Human imm
c	21	36.4	16.2	40388	4	ABL26342	Abl26342 Drosophil
c	22	36.2	16.1	5474	6	ABL33270	Abl33270 Human imm
	23	36	16.0	5153	2	AAT30347	Aat30347 Human YAP
	24	36	16.0	5153	9	ADD14716	Add14716 Human src
c	25	36	16.0	6725	6	ABL33208	Abl33208 Human imm
c	26	36	16.0	6725	6	ABL34554	Abl34554 Human met
c	27	35.8	15.9	441	8	ADB81862	Adb81862 Human cDN
c	28	35.8	15.9	502	6	ABQ88918	Abq88918 Human pro
	29	35.8	15.9	3002	4	ABL15414	Abl15414 Drosophil
c	30	35.8	15.9	7847	6	ABL34188	Abl34188 Human imm
c	31	35.6	15.8	6163	6	ABN80119	Abn80119 Human che
	32	35.4	15.7	932	4	AAK88641	Aak88641 Human dig
c	33	35.4	15.7	113515	6	ABL34175	Abl34175 Human imm
c	34	35.2	15.6	234	7	ABX54525	Abx54525 Bovine ES
	35	35.2	15.6	557	4	AAL18969	Aal18969 Human bre
c	36	35.2	15.6	8038	9	ADB85272	Adb85272 Mouse vit
c	37	35.2	15.6	11049	6	ABL32668	Abl32668 Human imm
c	38	35.2	15.6	11049	6	ABL92218	Abl92218 Chemicall
c	39	35.2	15.6	11049	6	ABL49321	Abl49321 Human pol
c	40	35.2	15.6	34319	8	ADA13460	Ada13460 Mouse rho
	41	35	15.6	332	4	AAK55869	Aak55869 Human imm
	42	35	15.6	1132	5	ABV28660	Abv28660 Human pro
	43	35	15.6	1132	5	ABV22830	Abv22830 Human pro
	44	35	15.6	2000	7	ADA71938	Ada71938 Rice gene
	45	35	15.6	3541	2	AAV07076	Aav07076 cDNA enco

ALIGNMENTS

RESULT 1
AAH23128
ID AAH23128 standard; DNA; 225 BP.
XX
AC AAH23128;
XX
DT 17-SEP-2001 (first entry)
XX
DE Osteoarthritis tissue-derived nucleic acid sequence #58.
XX
KW Osteoarthritis; infectious disorder; Crohn's disease; sepsis; human;
KW wound healing; osteopathic; anti-arthritic, anti-inflammatory; vulnerary;
KW antibacterial; antiallergic; ds.
XX
OS Homo sapiens.
XX
PN WO200153531-A2.
XX
PD 26-JUL-2001.
XX
PF 18-JAN-2001; 2001WO-US000016.
XX
PR 18-JAN-2000; 2000US-0176523P.
XX
PA (PHAA) PHARMACIA CORP.
XX
PI Phippard D, Vasanthakamur G, Dotson S, Ma X;
XX
DR WPI; 2001-451914/48.
XX
PT Substantially purified protein, polypeptide or their fragments, used to
PT identify a biologically active compound or composition and treat
PT mammalian osteoarthritis.
XX
PS Claim 1; Page 137; 144pp; English.
XX
CC Sequences AAH23071-23152 represent nucleic acid sequences derived from
CC osteoarthritis tissues. The sequences are useful as probes and for the
CC diagnosis or prognosis of mammalian osteoarthritis. The polynucleotides
CC and polypeptides of the invention are useful for generating diagnostic
CC reagents, as targets for small molecule drug development, generation of
CC therapeutics, and cloning genes. Specific antibodies are used to generate
CC enzyme linked immunosorbent assays for detection of osteoarthritis. The
CC invented molecules can be used to treat osteoarthritis or to analyse the
CC disease-modifying activity of osteoarthritis drugs. Other disorders
CC treatable using the nucleic acid sequences include atopic, inflammatory

```
CC   and infectious disorders e.g. Crohn's disease and sepsis, and wound
CC   healing
XX
SQ   Sequence 225 BP; 72 A; 38 C; 35 G; 80 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 225;  DB 4;  Length 225;
  Best Local Similarity   100.0%;  Pred. No. 1.1e-45;
  Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60

Qy         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120

Qy        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180

Qy        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225


RESULT 2
ACA04823
ID   ACA04823 standard; cDNA; 320 BP.
XX
AC   ACA04823;
XX
DT   28-MAY-2003  (first entry)
XX
DE   cDNA encoding human membrane associated protein fragment #271.
XX
KW   Human; ss; gene; microarray; membrane-associated protein; neuropathology;
KW   immunopathology; pancreatic disease; cancer; diabetes; hyperlipidaemia;
KW   pancreatic cholera; Alzheimer's disease; Huntington's disease; sarcoma;
KW   fibrocystic disease; leukaemia; adenocarcinoma; AIDS; allergy; anaemia;
KW   asthma; gout; dementia.
XX
OS   Homo sapiens.
XX
PN   US6492505-B1.
XX
PD   10-DEC-2002.
XX
PF   31-JAN-2000; 2000US-00495050.
XX
PR   01-FEB-1999;   99US-0118318P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Reddy R,  Guegler KJ,  Au-Young J;
XX
DR   WPI; 2003-327324/31.
XX
PT   Combination for research/diagnostic applications and for monitoring
PT   treatment of e.g., cancer, comprises polynucleotides comprising a
PT   fragment of gene encoding membrane-associated proteins, receptors or ion
PT   channels.
XX
PS   Claim 1; Col 251-252; 147pp; English.
XX
CC   The invention relates to a combination comprising several polynucleotide
CC   sequences comprising a fragment of gene encoding membrane-associated
CC   proteins, receptors or ion channels. The combination is useful as a
CC   probe, for research and diagnostic applications, for monitoring the
CC   expression of several expressed polynucleotides, in the diagnosis and
CC   monitoring of treatment of pancreatic disease, cancer, immunopathology or
CC   neuropathology, for investigating an individual's predisposition to the
CC   above disease, in genetic or gene expression analysis of polynucleotide
CC   sequences, to investigate cellular responses to infection or drug
CC   treatment, as hybridisable array elements in a microarray, to purify a
CC   subpopulation of mRNAs, cDNAs or genomic fragments in a sample, in
CC   diagnostics, prognostics and treatment regimens, in drug discovery and
CC   development, in toxicological and carcinogenicity studies, and in
CC   forensics or pharmacogenomics, to monitor the progression of disease, to
CC   monitor the efficacy of treatment, to diagnose the conditions of the
CC   pancreas e.g. diabetes, pancreatic cholera, hyperlipidaemia or
CC   fibrocystic disease, to diagnose a cancer e.g. leukaemia, adenocarcinoma
CC   or sarcoma, to diagnose immunopathologies e.g. AIDS, allergies, anaemia,
CC   asthma or gout, to diagnose neuropathologies e.g. Alzheimer's disease,
CC   dementia or Huntington's disease, to rapidly screen large numbers of
CC   candidate drug molecules and as query sequences against GenBank,
CC   SwissProt, BLOCKS and PRINTS databases. The combination is employed to
CC   fine tune the treatment regimen and thus the expression patterns
CC   associated with undesirable side effects are avoided. The present
CC   sequence represents a cDNA encoding a fragment of gene encoding human
CC   membrane-associated proteins, receptors or ion channels
XX
SQ   Sequence 320 BP; 105 A; 57 C; 53 G; 105 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 225;  DB 7;  Length 320;
  Best Local Similarity   100.0%;  Pred. No. 1.2e-45;
  Matches  225;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 TGATGGTAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCT 60

Qy         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ATATTGGTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTG 120

Qy        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCCCCCAAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCAT 180

Qy        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225
              |||||||||||||||||||||||||||||||||||||||||||||
Db        181 TTGATCTGCCTCCTCTGCTGAATCAATTAGGAATTTTTTTTTTTT 225


RESULT 3
AAV21209_01
Continuation (2 of 17) of AAV21209 from base 100001 (Methanococcus jannaschii circular
WP Sequence split into 17 fragments  LOCUS AAV21209 Accession Aav21209
WP      Fragment Name        Begin      End
WP      AAV21209_00              1    110000
WP      AAV21209_01         100001    210000
WP      AAV21209_02         200001    310000
WP      AAV21209_03         300001    410000
WP      AAV21209_04         400001    510000
WP      AAV21209_05         500001    610000
WP      AAV21209_06         600001    710000
WP      AAV21209_07         700001    810000
WP      AAV21209_08         800001    910000
WP      AAV21209_09         900001   1010000
WP      AAV21209_10        1000001   1110000
WP      AAV21209_11        1100001   1210000
WP      AAV21209_12        1200001   1310000
WP      AAV21209_13        1300001   1410000
WP      AAV21209_14        1400001   1510000
WP      AAV21209_15        1500001   1610000
WP      AAV21209_16        1600001   1664976

  Query Match             21.5%;  Score 48.4;  DB 2;  Length 110000;
  Best Local Similarity   51.9%;  Pred. No. 0.035;
  Matches  109;  Conservative    0;  Mismatches  101;  Indels    0;  Gaps    0;

Qy          7 TAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTG 66
              ||| |   | ||| | |   |||| |   | | |   ||| |||| | || || ||||
Db      51403 TAAATCTCTGCAGCCCTTCTATTTCCATCCATTATTATAGCTACATGTTTTGGTAAATTG 51462
```

```
Qy         67 GTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCC 126
              |||| |   ||     | |  |  |   |||    |||| ||| || ||  ||   |
Db      51463 TCTTTATCAATAGCCTCTTCTAAAAATCTTCTCGTAATTTTTAAAACTCCGGAGTTGTCT 51522

Qy        127 AAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATC 186
              ||||| | | | ||| |||  |     | ||||  |    |   |    |    ||| |||
Db      51523 AAAAATCTATAAAAATCAATTATTACTCTTTTTCCAATACTCTTTAATTTGTTTTTTATC 51582

Qy        187 TGCCTCCTCTGCTGAATCAATTAGGAATTT 216
              |  | |     |   | | ||||||||||
Db      51583 TTACCCAAAATCCCCACCTATTAGGAATTT 51612
```

RESULT 4
ABK28306/c
ID ABK28306 standard; DNA; 5518 BP.
XX
AC ABK28306;
XX
DT 23-APR-2002 (first entry)
XX
DE DNA transcription associated complementary genomic DNA #90.
XX
KW DNA transcription associated gene; peptide nucleic acid; PNA-oligomer;
KW PNA; cytosine methylation state; SNP; retroviral infection; gene; ds;
KW single nucleotide polymorphism; adenosine deaminase deficiency; cancer;
KW viral infection; Sezary syndrome; haematological disorder; tuberculosis;
KW immunological disorder; Werner syndrome; developmental disorder;
KW psoriasis; Rieger's syndrome; neurological disorder; erythropoiesis;
KW neurodegenerative disorder; Waardenburg syndrome; Niemann-Pick disease;
KW myelodysplastic syndrome; myocardial infarction; hypertension; arthritis;
KW angiogenesis; congenital heart disease; HDR syndrome; gene therapy;
KW polyglutamine disorder; solid tumour.
XX
OS Unidentified.
XX
PN WO200192565-A2.
XX
PD 06-DEC-2001.
XX
PF 06-APR-2001; 2001WO-EP003973.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-090046/12.
XX
PT New nucleic acids or oligomers, useful for diagnosing or treating
PT diseases associated with DNA transcription, e.g. immunological disorders,
PT Werner syndrome, psoriasis, myocardial infarction, solid tumors or
PT cancer.
XX
PS Claim 1; SEQ ID NO 180; 32pp; English.
XX
CC The invention relates to a nucleic acid, which comprises a segment of the
CC chemically pretreated DNA of genes associated with DNA transcription from
CC one of 346 sequences, and an oligomer, in particular an oligonucleotide
CC or peptide nucleic acid (PNA)-oligomer that hybridises to or is identical
CC to the chemically pretreated DNA of genes associated with DNA
CC transcription. The set of oligomer probes are useful for detecting the
CC cytosine methylation state and/or single nucleotide polymorphisms (SNPs)
CC in a chemically pretreated genomic DNA. The nucleic acids are useful for
CC diagnosing or treating diseases associated with DNA transcription
CC (particularly with the methylation status), e.g. adenosine deaminase
CC deficiency, viral infection, retroviral infection, Sezary syndrome,
CC haematological disorders, immunological disorders, Werner syndrome,
CC tuberculosis, developmental disorders, psoriasis, Rieger's syndrome,
CC neurological disorders, neurodegenerative disorders, Waardenburg
CC syndrome, Niemann-Pick disease, myelodysplastic syndrome, myocardial
CC infarction, hypertension, angiogenesis, erythropoiesis, congenital heart
CC disease, HDR syndrome, arthritis, polyglutamine disorders, solid tumours
CC or cancer. Sequences ABK28127-ABK28472 represent DNA transcription
CC associated genomic DNA molecules of the invention. Note: The sequence
CC data for this patent did not form part of the printed specification but
CC was obtained in electronic format directly from the European Patent
CC Office
XX
SQ Sequence 5518 BP; 1899 A; 52 C; 1031 G; 2536 T; 0 U; 0 Other;

Query Match 18.2%; Score 41; DB 6; Length 5518;
Best Local Similarity 51.4%; Pred. No. 1.5;
Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

```
Qy          7 TAAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTG 66
              ||  || ||| |     |||||||  | | || ||| |   |  | | ||  ||
Db       4088 TATATTTTTTTAAAACAAAAATTTTCATTTAACTATAAACTCCTCTAAAATATTAACGAA 4029

Qy         67 GTTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCC 126
              |  |  || | || | ||  ||   |   |||   | || ||| |||     |||
Db       4028 TTACTTAATTAATAATACTAAATATAATACTTCAAACTCTTAAAAACTAATTACTCCTAA 3969

Qy        127 AAAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATC 186
              || |||  | ||  || | ||||||  ||  |||| | | |  ||       | || | ||
Db       3968 AATAAAAAATACTTAATATAAAAACTACTATTTTTAAATTAAATACAACTTAATCTTATT 3909

Qy        187 TGCCT 191
              |  ||
Db       3908 TAACT 3904
```

RESULT 5
ABK31314/c
ID ABK31314 standard; DNA; 6809 BP.
XX
AC ABK31314;
XX
DT 23-APR-2002 (first entry)
XX
DE Signal transduction associated gene modified DNA #79.
XX
KW Human; signal transduction associated gene; cytosine methylation state;
KW CpG island; signal transduction associated disease; solid tumour; cancer;
KW antitumour; cytostatic; mutant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200926-A2.
XX
PD 03-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007472.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-147896/19.
XX
PT Oligonucleotide for diagnosis and therapy of diseases associated with
PT signal transduction e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with signal transduction.
XX
PS Claim 1; SEQ ID NO 157; 24pp; English.
XX
CC The present invention relates to chemically modified DNA sequences of
CC signal transduction associated genes. The DNA sequences are chemically

CC modified using a solution of bisulphite, hydrogen sulphite or disulphite.
CC Also disclosed are oligonucleotides and/or PNA oligomers for detecting
CC the cytosine methylation state (CpG islands) of these genes, and a method
CC for the diagnosis and/or therapy of genetic and epigenetic parameters of
CC genes associated with signal transduction. The genomic DNA can be
CC obtained from cells or cellular components which contain DNA, e.g. cell
CC lines, biopsies, blood, sputum, stool, urine, cerebral-spinal fluid,
CC tissue embedded in paraffin such as tissue from eyes, intestine, kidney,
CC brain, heart, prostate, lung, breast or liver, histologic object slides,
CC and all their possible combinations. The sequences of the invention are
CC useful for the diagnosis and therapy of diseases associated with signal
CC transduction e.g. solid tumours and cancer. ABK31158-ABK31545 represent
CC chemically pretreated genomic DNA sequences of different genes associated
CC with signal transduction, or their complementary sequences. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office
XX
SQ Sequence 6809 BP; 1792 A; 270 C; 1625 G; 3122 T; 0 U; 0 Other;

Query Match 17.8%; Score 40; DB 6; Length 6809;
Best Local Similarity 54.9%; Pred. No. 2.7;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy         17 CAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTT 76
              ||   ||||||    | ||| ||  | |  |  ||   ||| |   |||  ||| |   |
Db       1384 CACAAATAAAAACCAATATATATCGTAACACCACACTCTATTCATAATTTCTTTAATTAT 1325

Qy         77 CAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAA 136
                || | ||||||  || | ||  || | ||||   | || | | |  ||||  ||
Db       1324 ACAAAAACTATTTAAATTAAATCCCGTCTATCAATTTTTCGTTTTACTACAAATTACTTT 1265

Qy        137 ACAAAACAAAAAAACAACCTTTTT 160
               ||||||  ||  | ||  | |||
Db       1264 TCAAAACTTAATCATAAATTCTTT 1241
```

RESULT 6
ABL70557/c
ID ABL70557 standard; DNA; 6809 BP.
XX
AC ABL70557;
XX
DT 01-JUL-2002 (first entry)
XX
DE Chemically treated cell signalling DNA sequence#224.
XX
KW Cell signalling; cytosine methylation; cell signalling disease; cancer;
KW tumour; cytostatic; ds.
XX
OS Unidentified.
XX
PN WO200202807-A2.
XX
PD 10-JAN-2002.
XX
PF 29-JUN-2001; 2001WO-EP007471.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-154758/20.
XX
PT Nucleic acid, useful for diagnosis and therapy of diseases associated
PT with cell signaling e.g. cancer, comprises chemically modified genomic
PT sequences of genes associated with cell signaling.
XX
PS Claim 1; SEQ ID NO 447; 24pp + Sequence Listing; English.
XX

CC The invention relates to a nucleic acid comprising a sequence of at least
CC 18 bases of a segment of chemically pretreated DNA of genes associated
CC with cell signalling. The activity of the modified sequences of the
CC invention may be described as cytostatic. The object of the invention is
CC to provide the chemically modified DNA of genes associated with cell
CC signalling, as well as oligonucleotides and/or PNA-oligomers for
CC detecting cytosine methylations, as well as a method which is
CC particularly suitable for the diagnosis and/or therapy of genetic and
CC epigenetic parameters of genes associated with cell signalling. The
CC chemically modified DNA provided by the invention is useful for diagnosis
CC and therapy of diseases such as solid tumours and cancer. The sequences
CC given in records ABL70111-ABL70626 represent chemically pre-treated
CC genomic DNA's of genes associated with cell signalling. Note: The
CC sequence data for this patent is not represented in the printed
CC specification, but is based on sequence information supplied by the
CC European Patent Office
XX
SQ Sequence 6809 BP; 1792 A; 270 C; 1625 G; 3122 T; 0 U; 0 Other;

Query Match 17.8%; Score 40; DB 6; Length 6809;
Best Local Similarity 54.9%; Pred. No. 2.7;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

```
Qy         17 CAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTT 76
              ||   ||||||    | ||| ||  | |  |  ||   ||| |   |||  ||| |   |
Db       1384 CACAAATAAAAACCAATATATATCGTAACACCACACTCTATTCATAATTTCTTTAATTAT 1325

Qy         77 CAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAA 136
                || | ||||||  || | ||  || | ||||   | || | | |  ||||  ||
Db       1324 ACAAAAACTATTTAAATTAAATCCCGTCTATCAATTTTTCGTTTTACTACAAATTACTTT 1265

Qy        137 ACAAAACAAAAAAACAACCTTTTT 160
               ||||||  ||  | ||  | |||
Db       1264 TCAAAACTTAATCATAAATTCTTT 1241
```

RESULT 7
AAS61214/c
ID AAS61214 standard; DNA; 6809 BP.
XX
AC AAS61214;
XX
DT 29-JAN-2002 (first entry)
XX
DE Human gene regulation-associated gene oligonucleotide #169.
XX
KW Human; Gene regulation-associated gene; severe combined immunodeficiency;
KW cardiac damage; inflammatory response; Haemophilia; Werner syndrome;
KW asthma; HDR syndrome; congenital heart defect; Saethre-Chotzen syndrome;
KW renal disease; Preeclampsia; cardiac allograft vascular disease;
KW colorectal cancer; thyroid cancer; oesophageal cancer; ds; tumour;
KW immunostimulant; cardiant; antiinflammatory; coagulant; antiasthmatic;
KW nephrotropic; gynecological; anti-tumour; immunosuppressive; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200177375-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001WO-EP003968.
XX
PR 06-APR-2000; 2000DE-01019058.
PR 07-APR-2000; 2000DE-01019173.
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-017470/02.
XX

PT New nucleic acid sequences from chemically modified genes associated with
PT gene regulation, useful for analyzing cytosine methylations for diagnosis
PT and therapy of diseases e.g. severe combined immunodeficiency disease.
XX
PS Claim 1; SEQ ID NO 173; 26pp; English.
XX
CC The invention relates to 224 nucleic acid sequences comprising at least
CC 18 bases of a chemically pretreated gene associated with gene regulation
CC selected from 43 known genes (or complementary sequences). The chemical
CC pretreatment converts cytosine bases unmethylated at the 5-position to
CC uracil or another base with hybridisation behaviour dissimilar to
CC cytosine, to enable analysis of cytosine methylations. The DNA sequences,
CC oligomers (or sets/arrays) and method are useful in the diagnosis of
CC diseases (or predisposition to diseases) associated with gene regulation
CC and in therapy of such diseases, by enabling analysis of the cytosine
CC methylation patterns of such genes, kits are provided. They are
CC especially useful in diagnosis and therapy of e.g. severe combined
CC immunodeficiency disease, cardiac disorders, haemophilia, solid tumours
CC and cancer, Werner syndrome, asthma, HDR syndrome, Saethre-Chotzen
CC syndrome, renal disease, preeclampsia, graft versus-host disease. The
CC present sequence is a sequence included in the sequence data for this
CC specification and is associated with the human gene regulation-associated
CC genes. Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 6809 BP; 1792 A; 270 C; 1625 G; 3122 T; 0 U; 0 Other;

Query Match 17.8%; Score 40; DB 6; Length 6809;
Best Local Similarity 54.9%; Pred. No. 2.7;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 17 CAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTT 76
Db 1384 CACAAATAAAAACCAATATATATCGTAACACCACACTCTATTCATAATTTCTTTAATTAT 1325

Qy 77 CAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAA 136
Db 1324 ACAAAAACTATTTAAATTAAATCCCGTCTATCAATTTTTCGTTTTACTACAAATTACTTT 1265

Qy 137 ACAAAACAAAAAAACAACCTTTTT 160
Db 1264 TCAAAACTTAATCATAAATTCTTT 1241

RESULT 8
ABN80174/c
ID ABN80174 standard; DNA; 6809 BP.
XX
AC ABN80174;
XX
DT 15-JUL-2002 (first entry)
XX
DE Human chemically modified disease associated gene SEQ ID NO 191.
XX
KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;
KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;
KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;
KW antidiabetic; cytostatic; anticonvulsant; ds.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200200927-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007536.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130908/17.
XX
PT Novel nucleic acid useful for diagnosis and therapy of diseases
PT associated with development genes such as diabetes, comprises a sequence
PT of a segment of chemically pretreated DNA of genes associated with
PT development.
XX
PS Claim 1; SEQ ID NO 191; 27pp; English.
XX
CC The invention relates to a nucleic acid (I) comprising a sequence at
CC least 18 bases in length of a segment of chemically pretreated DNA (II)
CC of genes associated with development selected from 87 genes listed in the
CC specification such as ACCPN, ADFN, or AFD1 and comprising one of 350
CC sequences (ABN79984-ABN80333) or their complements. The invention is
CC useful for the diagnosis or therapy of diseases associated with
CC development genes, in particular disease related to homeobox containing
CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes
CC associated with congenital heart disease, epilepsy, diseases related to
CC histone deacetylation, Currarino syndrome, diseases related with the
CC development of the brain and limb girdle muscular dystrophy and dwarfism.
CC Oligomers specific to each of the genes are useful for detecting the
CC methylation state of all CpG dinucleotides within the 350 sequences or
CC (II) and their complementary sequences, as primer oligonucleotides for
CC the amplification of the 350 sequences, (II) and/or their complements and
CC as oligomer probes for detecting the cytosine methylation state and/or
CC single nucleotide polymorphisms (SNPs). Note: The sequence data for this
CC patent did not form part of the printed specification but is based on
CC sequence information supplied to Derwent by the European Patent Office
XX
SQ Sequence 6809 BP; 1792 A; 270 C; 1625 G; 3122 T; 0 U; 0 Other;

Query Match 17.8%; Score 40; DB 6; Length 6809;
Best Local Similarity 54.9%; Pred. No. 2.7;
Matches 79; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

Qy 17 CAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTT 76
Db 1384 CACAAATAAAAACCAATATATATCGTAACACCACACTCTATTCATAATTTCTTTAATTAT 1325

Qy 77 CAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAA 136
Db 1324 ACAAAAACTATTTAAATTAAATCCCGTCTATCAATTTTTCGTTTTACTACAAATTACTTT 1265

Qy 137 ACAAAACAAAAAAACAACCTTTTT 160
Db 1264 TCAAAACTTAATCATAAATTCTTT 1241

RESULT 9
ABA92787_1/c
Continuation (2 of 7) of ABA92787 from base 100001 (Buchnera sp. genomic DNA SEQ ID NO:1
WP Sequence split into 7 fragments LOCUS ABA92787 Accession Aba92787

WP	Fragment Name	Begin	End
WP	ABA92787_0	1	110000
WP	ABA92787_1	100001	210000
WP	ABA92787_2	200001	310000
WP	ABA92787_3	300001	410000
WP	ABA92787_4	400001	510000
WP	ABA92787_5	500001	610000
WP	ABA92787_6	600001	640681

Query Match 17.0%; Score 38.2; DB 6; Length 110000;
Best Local Similarity 55.7%; Pred. No. 11;
Matches 73; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

Qy 32 AAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAGAATATTTAG 91
Db 98302 AATCAAATTAATAAACTAGTATATTTTCAATTATGACTTTTCAATAATTTTATTATTAAG 98243

Qy 92 TTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAAAAC 151

```
Db      98242 TAATAATTGTTATTTATAATAATACTTTCATCTAAAATAAACAATACAAAAAAAAAAAAA 98183

Qy        152 AACCTTTTTAA 162
              ||  ||||| |
Db      98182 AAGATTTTTTA 98172
```

RESULT 10
ABL33263/c
ID ABL33263 standard; DNA; 12393 BP.
XX
AC ABL33263;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human immune system associated gene SEQ ID NO: 1236.
XX
KW Human; immune system disease; cytosine methylation; antiasthmatic;
KW antiarteriosclerotic; antianaemic; cytostatic; nootropic;
KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;
KW ds.
XX
OS Homo sapiens.
XX
PN WO200200928-A2.
XX
PD 03-JAN-2002.
XX
PF 02-JUL-2001; 2001WO-EP007537.
XX
PR 30-JUN-2000; 2000DE-01032529.
PR 01-SEP-2000; 2000DE-01043826.
XX
PA (EPIG-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-130909/17.
XX
PT Nucleic acid comprising fragment of chemically modified gene, useful for
PT diagnosis and treatment of diseases associated with abnormal cytosine
PT methylation.
XX
PS Claim 1; SEQ ID NO 1236; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
CC diseases. The present sequence is a gene of the invention
XX
SQ Sequence 12393 BP; 3484 A; 219 C; 2406 G; 6282 T; 0 U; 2 Other;

  Query Match   16.9%; Score 38; DB 6; Length 12393;
  Best Local Similarity 49.5%; Pred. No. 9;
  Matches   98; Conservative   0; Mismatches 100; Indels   0; Gaps   0;

```
Qy         22 ATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAA 81
              | |||| || ||| |||  ||        || || |   |||    |||| |||| | ||
Db       6387 AAAAAACTTAAAAAAAAAAATTTTAAAAAATAATTTAAAATAAAACTTTTTCCTTTAAAA 6328

Qy         82 GAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAA 141
               ||   | | |||   | | | |||| |||   |     |  | | | |||  ||| |||
Db       6327 AAAACATAATTTTATTTAACGATTTTTAAAAAAATCAAATAACTCGATAAATTAAATAAA 6268
```

```
Qy        142 ACAAAAAAACAACCTTTTTAAGAGTTGATGGCTACTCATTTGATCTGCCTCCTCTGCTGA 201
              || ||||       | |    | ||       | ||    | |||  |   |   || |
Db       6267 TAAATAAAAATCAAATATACTCAATTCTAACTAATTCTCCCGCTCTCTCAAATAACCTTA 6208

Qy        202 ATCAATTAGGAATTTTTT 219
              | ||| |    ||  |||
Db       6207 AACAAATTTCTATCATTT 6190
```

RESULT 11
ADA03026/c
ID ADA03026 standard; DNA; 96588 BP.
XX
AC ADA03026;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human MBNL carcinoma associated gene, SEQ ID NO:1544.
XX
KW Human; carcinoma associated; oncogene; carcinoma; cancer; breast;
KW prostate; lymphoma; leukaemia; cytostatic; gene therapy; drug screening;
KW gene; ds.
XX
OS Homo sapiens.
XX
PN WO2003057146-A2.
XX
PD 17-JUL-2003.
XX
PF 26-DEC-2002; 2002WO-US041414.
XX
PR 26-DEC-2001; 2001US-00035832.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-587068/55.
XX
PT New recombinant nucleic acid encoding carcinoma associated protein,
PT useful for preparing compositions for treating carcinomas.
XX
PS Claim 1; SEQ ID NO 1544; 245pp; English.
XX
CC The invention relates to recombinant carcinoma associated (CA) nucleic
CC acid sequences from mouse and human (ADA01482-ADA03094), and to
CC recombinant carcinoma associated proteins (CAP) encoded by them. The
CC invention also encompasses expression vectors and host cells comprising a
CC CA nucleic acid, a polypeptide (especially an antibody) that specifically
CC binds to the protein, and a biochip comprising CA nucleic acid or
CC fragments thereof. The sequences of the invention were identified using
CC oncogenic retroviruses, which insert into the genome of the host organism
CC at random. Many of these do not carry transduced host oncogenes or
CC pathogenic trans-acting viral genes, meaning that cancer incidence is a
CC direct consequence of the effects of proviral integration into host
CC protooncogenes. The CA nucleic acid sequences can be used to diagnose
CC carcinoma (especially breast cancer, prostate cancer, lymphoma or
CC leukaemia) or a propensity to carcinoma by determination of the sequence
CC of a CA gene, or by determination of CA gene expression in particular
CC tissues. CA nucleic acids, proteins and antibodies are also useful as
CC therapeutic agents and in screening and evaluating drug candidates. The
CC present sequence represents a specifically claimed human CA nucleic acid
CC sequence of the invention. Note: The complete sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 96588 BP; 29654 A; 16428 C; 18069 G; 32437 T; 0 U; 0 Other;

  Query Match   16.8%; Score 37.8; DB 8; Length 96588;
  Best Local Similarity 51.5%; Pred. No. 13;
  Matches   87; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

```
Qy          8 AAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGG 67
              | ||||      | |||   | | | |    | ||     ||||||  ||  |||
Db      61750 ATGTTGAGCATTGTATATTTTATTATAATTTAATTGTCTTCCATGTCATTTTATTAGTA 61691

Qy         68 TTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCA 127
                ||   |    |  |  || ||    ||  ||  |||||  |   |  | |||||  |
Db      61690 AATTTTATAATAAGCATAATGTATACACATGCATTTTTTTGGAGAGTCTGTTGTCCAGCC 61631

Qy        128 AAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTAC 176
              |||||  ||| |||| ||||||| || |  | ||  | || |  || ||
Db      61630 AAAAAAAAAAAAAAAAAAAAAAAAAAACAGTATAGAACTTTAATGCCAC 61582
```

RESULT 12
ADB72764/c
ID ADB72764 standard; DNA; 96588 BP.
XX
AC ADB72764;
XX
DT 04-DEC-2003 (first entry)
XX
DE Human MBNL gene.
XX
KW human; ds; cytostatic; gene therapy; vaccine; carcinoma; lymphomas;
KW cancer; neoplasm; adenocarcinoma; sarcoma; gene.
XX
OS Homo sapiens.
XX
PN WO2003008583-A2.
XX
PD 30-JAN-2003.
XX
PF 26-DEC-2001; 2001WO-US051291.
XX
PR 02-MAR-2001; 2001US-00798586.
PR 23-OCT-2001; 2001US-00004113.
PR 08-NOV-2001; 2001US-00052482.
PR 30-NOV-2001; 2001US-00997722.
PR 20-DEC-2001; 2001US-00034650.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-239337/23.
XX
PT New recombinant nucleic acid, useful for treating carcinomas, lymphomas,
PT cancers, neoplasm, adenocarcinoma, or sarcomas.
XX
PS Claim 1; SEQ ID NO 592; 2304pp; English.
XX
CC The invention relates to a novel recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the 660 sequences fully defined
CC in the specification. A polynucleotide of the invention has cytostatic
CC activity, and may have a use in gene therapy, or in a vaccine. The
CC recombinant nucleic acids and polypeptides are useful for treating
CC carcinomas, e.g. lymphomas, cancers, neoplasm, adenocarcinoma, and
CC sarcomas. The present sequence represents a human gene of the invention.
XX
SQ Sequence 96588 BP; 29654 A; 16428 C; 18069 G; 32437 T; 0 U; 0 Other;

  Query Match   16.8%; Score 37.8; DB 9; Length 96588;
  Best Local Similarity 51.5%; Pred. No. 13;
  Matches   87; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

```
Qy          8 AAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGG 67
              | ||||      | |||   | | | |    | ||     ||||||  ||  |||
Db      61750 ATGTTGAGCATTGTATATTTTATTATAATTTAATTGTCTTCCATGTCATTTTATTAGTA 61691

Qy         68 TTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCA 127
                ||   |    |  |  || ||    ||  ||  |||||  |   |  | |||||  |
Db      61690 AATTTTATAATAAGCATAATGTATACACATGCATTTTTTTGGAGAGTCTGTTGTCCAGCC 61631

Qy        128 AAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTAC 176
              |||||  ||| |||| ||||||| || |  | ||  | || |  || ||
Db      61630 AAAAAAAAAAAAAAAAAAAAAAAAAAACAGTATAGAACTTTAATGCCAC 61582
```

RESULT 13
ADC85506/c
ID ADC85506 standard; DNA; 96588 BP.
XX
AC ADC85506;
XX
DT 01-JAN-2004 (first entry)
XX
DE Human Mbnl genomic sequence.
XX
KW Cytostatic; gene therapy; vaccine; cancer; carcinoma-associated gene; CA;
KW secreted; transmembrane; intracellular; ds.
XX
OS Homo sapiens.
XX
PN WO2003045230-A2.
XX
PD 05-JUN-2003.
XX
PF 02-DEC-2002; 2002WO-US038582.
XX
PR 30-NOV-2001; 2001US-00997722.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW, Engelhard EK;
XX
DR WPI; 2003-513603/48.
XX
PT New recombinant nucleic acid comprising a nucleotide sequence of any of
PT the carcinoma-associated (CA) genes, useful for screening for drug
PT candidates for diagnosing or treating carcinomas.
XX
PS Claim 1; SEQ ID NO 292; 983pp; English.
XX
CC The invention relates to a recombinant nucleic acid comprising a
CC nucleotide sequence selected from any of the fully defined carcinoma-
CC associated (CA) genes from the 50 tables given in the specification. The
CC CA proteins are secreted, transmembrane or intracellular proteins. The
CC recombinant nucleic acids are useful for screening for drug candidates
CC for diagnosing or treating carcinomas. Sequences given in ADC85215-
CC ADC85514 represent CA genes of the invention.
XX
SQ Sequence 96588 BP; 29654 A; 16428 C; 18069 G; 32437 T; 0 U; 0 Other;

  Query Match   16.8%; Score 37.8; DB 9; Length 96588;
  Best Local Similarity 51.5%; Pred. No. 13;
  Matches   87; Conservative   0; Mismatches   82; Indels   0; Gaps   0;

```
Qy          8 AAGTTGTTTCAGGCATAAAATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGG 67
              | ||||      | |||   | | | |    | ||     ||||||  ||  |||
Db      61750 ATGTTGAGCATTGTATATTTTATTATAATTTAATTGTCTTCCATGTCATTTTATTAGTA 61691

Qy         68 TTTTACCTTCAGAAGAATATTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCA 127
                ||   |    |  |  || ||    ||  ||  |||||  |   |  | |||||  |
Db      61690 AATTTTATAATAAGCATAATGTATACACATGCATTTTTTTGGAGAGTCTGTTGTCCAGCC 61631

Qy        128 AAAAACGAAACAAAACAAAAAAACAACCTTTTTAAGAGTTGATGGCTAC 176
              |||||  ||| |||| ||||||| || |  | ||  | || |  || ||
Db      61630 AAAAAAAAAAAAAAAAAAAAAAAAAAACAGTATAGAACTTTAATGCCAC 61582
```

RESULT 14
AAX05715/c
ID AAX05715 standard; DNA; 2270 BP.
XX
AC AAX05715;
XX

```
DT   07-MAY-1999  (first entry)
XX
DE   Human protein phosphatase (PROPHO) encoding DNA.
XX
KW   Protein phosphatase; PROPHO; apoptosis; AIDS; Alzheimer's Disease;
KW   Acquired Immune Deficiency Syndrome; Parkinson's Disease; inflammation;
KW   cell proliferation; Addison's disease; allergy; anemia; cancer; bone;
KW   leukemia; breast; brain; human; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             70. .1509
FT                   /*tag=  a
FT                   /product= "protein phosphatase (PROPHO)"
XX
PN   WO9856925-A1.
XX
PD   17-DEC-1998.
XX
PF   11-JUN-1998;   98WO-US011614.
XX
PR   11-JUN-1997;   97US-00873093.
XX
PA   (INCY-) INCYTE PHARM INC.
XX
PI   Bandman O,  Goli SK,  Lal P,  Corley NC,  Zhang H;
XX
DR   WPI; 1999-080906/07.
DR   P-PSDB; AAW94283.
XX
PT   New substantially purified human protein phosphatase (PROPHO) - useful in
PT   the diagnosis, prevention or treatment of inflammation, cancer, and
PT   disorders associated with apoptosis.
XX
PS   Claim 5; Fig 1A-G; 73pp; English.
XX
CC   This DNA encodes a human protein phosphatase (PROPHO). Host cells
CC   containing a vector comprising the PROPHO nucleic acid are used for the
CC   recombinant production of the protein. PROPHO forms a composition in the
CC   treatment or prevention of apoptosis-related disorders (e.g. Acquired
CC   Immune Deficiency Syndrome (AIDS), Alzheimer's Disease and Parkinson's
CC   Disease), and in the stimulation of cell proliferation. Antagonists of
CC   the protein are useful in treating inflammation (e.g. Addison's disease,
CC   allergies and anemia), and disorders associated with cell proliferation
CC   (including various cancers like leukemia, and cancers affecting bone,
CC   breast and brain). Complementary polynucleotides are useful in detecting
CC   polynucleotides that encode PROPHO, useful in the diagnosis of conditions
CC   associated with the expression of PROPHO, and in assays that detect
CC   activation or induction of various cancers. PROPHO is useful in producing
CC   antibodies or screening libraries of pharmaceutical agents in order to
CC   identify those that bind to PROPHO
XX
SQ   Sequence 2270 BP; 693 A; 344 C; 515 G; 716 T; 0 U; 2 Other;

  Query Match             16.6%;  Score 37.4;  DB 2;  Length 2270;
  Best Local Similarity   53.0%;  Pred. No. 10;
  Matches   80;  Conservative    0;  Mismatches   71;  Indels    0;  Gaps    0;

Qy         29 TTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAGAATATT 88
              |||| ||| ||| ||| |      | || ||||||  |  | |||   |  |||     |
Db       2197 TTGACATACATTTTGATGTGTGTAGTTAGGCTATACAGAATATACAAACCAAAGCTGTAT 2138

Qy         89 TAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAAAA 148
              | |   ||  ||||||||||||  |    | ||    | || ||  |  ||   ||||
Db       2137 GAATAATACATAGGTTTTTCAAAAGTTATGTTTCATAAAGAATACAGAGAAAGGAAAAAT 2078

Qy        149 AACAACCTTTTTAAGAGTTGATGGCTACTCA 179
              | | ||   |  || |  | |   || || |
Db       2077 AGCCACACATCCAATATCTCACAACTTCTAA 2047

RESULT 15
AAL21862/c
ID   AAL21862 standard; cDNA; 778 BP.
XX
AC   AAL21862;
XX
DT   07-DEC-2001  (first entry)
XX
DE   Human breast cancer expressed polynucleotide 14319.
XX
KW   Human; breast cancer; cell marker; cytostatic; ss.
XX
OS   Homo sapiens.
XX
PN   WO200151628-A2.
XX
PD   19-JUL-2001.
XX
PF   10-JAN-2001; 2001WO-US000798.
XX
PR   14-JAN-2000; 2000US-0176077P.
PR   14-MAR-2000; 2000US-0189167P.
PR   24-MAR-2000; 2000US-0192099P.
PR   29-MAR-2000; 2000US-0193480P.
PR   15-MAY-2000; 2000US-0205230P.
PR   09-JUN-2000; 2000US-0211315P.
PR   25-JUL-2000; 2000US-0220534P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Lillie J,  Xu Y,  Wang Y,  Steinmann K;
XX
DR   WPI; 2001-451856/48.
XX
PT   New peptide useful as a marker for the diagnosis of breast cancer.
XX
PS   Claim 1; Page 2564-2565; 3695pp; English.
XX
CC   The invention relates to human breast cancer expressed polynucleotides
CC   (AAL07544-AAL26789) and methods of assessing whether a patient is
CC   afflicted with breast cancer by examining the correlation between the
CC   expression of certain markers and the cancerous state of breast cells.
CC   The polynucleotides and encoded polypeptides are potential markers for
CC   detecting, diagnosing, monitoring, characterising treating and
CC   potentially preventing breast cancer. The polynucleotides and encoded
CC   polypeptides are also useful for isolating compounds with cytostatic
CC   activity
XX
SQ   Sequence 778 BP; 220 A; 134 C; 180 G; 244 T; 0 U; 0 Other;

  Query Match             16.5%;  Score 37.2;  DB 4;  Length 778;
  Best Local Similarity   55.4%;  Pred. No. 9.8;
  Matches   72;  Conservative    0;  Mismatches   58;  Indels    0;  Gaps    0;

Qy         27 ATTTGAAATAAATTATGAGGCTCCATGATATGCTATATTGGTTTTACCTTCAGAAGAATA 86
              | || |  || |||       ||| || | || |||  |  |    ||  |  | |||||
Db        136 AATTTAGCTATATTTATCTTCTCAATCAAATTCTACTTCAGAGGTAAGTAAAAAAGAAGT 77

Qy         87 TTTAGTTTCACTCAGGTTTTTCAAAGCTACGCTGTCCCCCAAAAAACGAAACAAAACAAA 146
              | |||     | ||     |  | |  | | | ||||  ||||||  ||| |||| |||
Db         76 TATAGCAAATCACATAAAATGGACAAAAAAGGTATCCCAAAAAAAAAAAAAAAAAAAAAA 17

Qy        147 AAAACAACCT 156
              ||||  ||||
Db         16 AAAAGTACCT 7


Search completed: July 27, 2004, 18:19:56
Job time : 411 secs
```